

09/475704

146 938

STIC-Biotech/ChemLib

1179

From: Whiteman, Brian  
Sent: Monday, March 07, 2005 7:20 AM  
To: STIC-Biotech/ChemLib  
Subject: seq search

09/475,704 Barnett et al. 12/30/99

SEQ ID NO: 3 and 4

Please search protein and NA databases

- 1) default
- 2) interference

Thank you,

Brian Whiteman  
Remsen, 2D14  
mail box 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
(571) 272-0764

RECEIVED  
MAR - 7 2005  
STIC

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2- \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence :# \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 21:18:31 ; Search time 916.106 Seconds

(without alignments)  
9798.345 Million cell updates/sec

Title: US-09-475-704A-4

Perfect score: 1509  
Sequence: 1 atggcgccgccgcagcatc.....gcggcccttgagcagtraa 1509

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5537552 seqs, 2974263231 residues

Total number of hits satisfying chosen parameters: 11075104

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1509	100.0	1509	US-09-899-575-4	Sequence 4, Appl
2	1507.4	99.9	1509	US-09-967-464-64	Sequence 64, Appl
3	1504.2	99.7	1509	US-09-967-464-68	Sequence 68, Appl
4	1489.8	98.7	1509	US-09-899-575-21	Sequence 21, Appl
5	1297.6	86.0	1494	US-09-899-575-51	Sequence 51, Appl
6	1294.8	85.8	1491	US-09-899-575-99	Sequence 99, Appl
7	1276.8	84.6	1479	US-09-967-464-63	Sequence 63, Appl
8	1276.8	84.6	1479	US-09-899-575-3	Sequence 3, Appl
9	1275.2	84.5	3462	US-10-190-435-18	Sequence 18, Appl
10	1275.2	84.5	3462	US-10-190-435-16	Sequence 16, Appl
11	1275.2	84.5	4419	US-10-190-435-19	Sequence 19, Appl

12	1275.2	84.5	4419	US-10-190-305A-14	Sequence 14, Appl
13	1275.2	84.5	4406	US-10-190-435-35	Sequence 35, Appl
14	1275.2	84.5	4606	US-10-190-435-34	Sequence 34, Appl
15	1275.2	84.5	4615	US-10-190-435-36	Sequence 36, Appl
16	1275.2	84.5	4716	US-10-190-435-38	Sequence 38, Appl
17	1275.2	84.5	4716	US-10-190-435-17	Sequence 17, Appl
18	1275.2	84.5	4716	US-10-190-305A-13	Sequence 13, Appl
19	1273.6	84.4	2742	US-10-190-435-20	Sequence 20, Appl
20	1273.6	84.4	2742	US-10-190-305A-15	Sequence 15, Appl
21	1273.6	84.4	3330	US-10-190-435-9	Sequence 9, Appl
22	1273.6	84.4	3330	US-10-190-435-10	Sequence 10, Appl
23	1273.6	84.4	3330	US-10-190-435-11	Sequence 11, Appl
24	1273.6	84.4	5145	US-10-190-435-12	Sequence 12, Appl
25	1273.6	84.4	5145	US-10-190-305A-12	Sequence 59, Appl
26	1272.6	84.3	4713	US-10-190-435-59	Sequence 59, Appl
27	1272.6	84.3	4713	US-10-190-305A-83	Sequence 83, Appl
28	1271.6	84.3	5184	US-10-190-435-58	Sequence 58, Appl
29	1271.6	84.3	5184	US-10-190-305A-82	Sequence 82, Appl
30	1271.2	84.2	1479	US-09-967-464-67	Sequence 67, Appl
31	1270.2	84.2	2742	US-10-190-435-57	Sequence 57, Appl
32	1270.2	84.2	2742	US-10-190-305A-81	Sequence 81, Appl
33	1260.8	83.6	1479	US-09-899-575-20	Sequence 20, Appl
34	1260.8	83.6	1515	US-10-387-336-4	Sequence 4, Appl
35	1201	79.6	4472	US-10-387-336-75	Sequence 75, Appl
36	1201	79.6	4608	US-10-387-336-76	Sequence 76, Appl
37	1201	79.6	4689	US-10-387-336-74	Sequence 74, Appl
38	1201	79.6	4766	US-10-387-336-73	Sequence 73, Appl
39	1199.4	79.5	2799	US-10-241-009-18	Sequence 18, Appl
40	1199.4	79.5	2799	US-10-190-434B-18	Sequence 18, Appl
41	1199.4	79.5	2799	US-10-190-305A-11	Sequence 11, Appl
42	1199.4	79.5	3305	US-10-241-009-17	Sequence 17, Appl
43	1199.4	79.5	3305	US-10-190-434B-17	Sequence 17, Appl
44	1199.4	79.5	3496	US-10-241-009-15	Sequence 15, Appl
45	1199.4	79.5	3496	US-10-190-434B-15	Sequence 15, Appl

## ALIGNMENTS

RESULT 1  
US-09-899-575-4  
; Sequence 4, Application US/09899575  
; Publication No. US20030223961A1  
GENERAL INFORMATION:  
; APPLICANT: Zur Megele, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Estrelita Janse  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
FILE REFERENCE: P01631.102  
CURRENT APPLICATION NUMBER: US/09/899,575  
PRIOR FILING DATE: 2001-07-05  
PRIOR APPLICATION NUMBER: 09/475,704  
NUMBER OF SEQ ID NOS: 135  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4  
LENGTH: 1509  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag  
; OTHER INFORMATION: of HIV strain AFI10967  
US-09-899-575-4  
Query Match 100.0%; Score 1509; DB 10; Length 1509;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 ATGGCGCCGCCGCCAGCATCTGCGCGCGGCGAAGACTGGAAGTGGGAGATCCGC 60  
1 ATGGCGCCGCCGCCAGCATCTGCGCGCGGCGAAGACTGGAAGTGGGAGATCCGC 60

```
QY 61 CTGCGCCCGCGGCAAGACCTTACATGCTGAAGACCTGTGTGGCCACGCCGAG 120
Db 61 CTGCGCCCGCGGCAAGACCTTACATGCTGAAGACCTGTGTGGCCACGCCGAG 120
QY 121 CTGGAAGGGCTTGCCTGTGAACCCCGGCTGTGAGAGACCGCGAGGGCTGTGAAGAGATC 180
Db 121 CTGGAAGGGCTTGCCTGTGAACCCCGGCTGTGAGAGACCGCGAGGGCTGTGAAGAGATC 180
QY 181 ATGAAGACAGCTGAGCCGCTGTGACAGACCGGACCCGAGAGCTGTGCAAC 240
Db 181 ATGAAGACAGCTGAGCCGCTGTGACAGACCGGACCCGAGAGCTGTGCAAC 240
QY 241 ACCGTTGACCCCTGTACTGCGTGAACCGCGGATGAGTGTCCGAGACCAAGAGGTC 300
Db 241 ACCGTTGACCCCTGTACTGCGTGAACCGCGGATGAGTGTCCGAGACCAAGAGGTC 300
QY 301 CTGGAACAAGATGAGAGAGAGAGACCAAGTCCAGACAGAACCCAGAGGCCAAGAG 360
Db 301 CTGGAACAAGATGAGAGAGAGAGACCAAGTCCAGACAGAACCCAGAGGCCAAGAG 360
QY 361 GCCGACGGCAGAGGTGAGCCAGAACTACCCATGTGCAAGAACCTGACAGGGCAGATGTTG 420
Db 361 GCCGACGGCAGAGGTGAGCCAGAACTACCCATGTGCAAGAACCTGACAGGGCAGATGTTG 420
QY 421 CACCAAGGCGCATGAGCCCGGCAACCTGGAACGCGTGGTGAAGTGTGAGAGAGAGGTC 480
Db 421 CACCAAGGCGCATGAGCCCGGCAACCTGGAACGCGTGGTGAAGTGTGAGAGAGAGGTC 480
QY 481 TTCAAGCCCGGAGGTATCCCATGTTCAACCGCCCTGAGAGAGAGGCGCACCCCGCAGAG 540
Db 481 TTCAAGCCCGGAGGTATCCCATGTTCAACCGCCCTGAGAGAGAGGCGCACCCCGCAGAG 540
QY 541 CTGAACAGATTTTGAACACCGTGGCGGCGCACAGGCGCGCATGTGTAAGAGAC 600
Db 541 CTGAACAGATTTTGAACACCGTGGCGGCGCACAGGCGCGCATGTGTAAGAGAC 600
QY 601 ACCATCAAGAGAGAGCGCGGAGTGGAGCGGCTGACCCCGTGCAGGCGGCGGCGGCGGTC 660
Db 601 ACCATCAAGAGAGAGCGCGGAGTGGAGCGGCTGACCCCGTGCAGGCGGCGGCGGCGGTC 660
QY 661 GCCCGGCGGAGATGCGGACCCCGCGGAGAGCATGCGCGGCGCACGACCCCTG 720
Db 661 GCCCGGCGGAGATGCGGACCCCGCGGAGAGCATGCGCGGCGCACGACCCCTG 720
QY 721 CAGGAGCATGCTGTGATGACCAAGACCCCGGCTGTGGGAGCATTTACAG 780
Db 721 CAGGAGCATGCTGTGATGACCAAGACCCCGGCTGTGGGAGCATTTACAG 780
QY 781 CGGTGATCATCTGTGGCTGTGAACAAGATCGTGGATGTACAGCCCGTGAAGCATCTG 840
Db 781 CGGTGATCATCTGTGGCTGTGAACAAGATCGTGGATGTACAGCCCGTGAAGCATCTG 840
QY 841 GACATCCGCGAGAGGCCCAAGAGACCTTCCGCGACCTACGTGAGCCGCTTTCTTAAGAC 900
Db 841 GACATCCGCGAGAGGCCCAAGAGACCTTCCGCGACCTACGTGAGCCGCTTTCTTAAGAC 900
QY 901 CTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Db 901 CTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 961 CAGAAAGCAGAACCCCGACTGCAAGACCAATCTGCGCGCTCTGCGGCCCGGCGCACCTTG 1020
Db 961 CAGAAAGCAGAACCCCGACTGCAAGACCAATCTGCGCGCTCTGCGGCCCGGCGCACCTTG 1020
QY 1021 GAGGAGATGATGACCGCTGCGAGGGCGTGGCGGCGCCCGGACAGAGGCCGCGTGTG 1080
Db 1021 GAGGAGATGATGACCGCTGCGAGGGCGTGGCGGCGCCCGGACAGAGGCCGCGTGTG 1080
QY 1081 GCCGAGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 GCCGAGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
```

```
QY 1141 GGCCCCCGGCGAAGTCAAGTCTTCAACTGTGCGCAAGAGAGGCCACATGCGCAAGAC 1200
Db 1141 GGCCCCCGGCGAAGTCAAGTCTTCAACTGTGCGCAAGAGAGGCCACATGCGCAAGAC 1200
QY 1201 TGCGCGCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 TGCGCGCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 GACTGACCGAGAGCGCAGAGCCAACTTCTGTGGCAAGATCTGAGCCAGCCAGAGAGAGAG 1320
Db 1261 GACTGACCGAGAGCGCAGAGCCAACTTCTGTGGCAAGATCTGAGCCAGCCAGAGAGAGAG 1320
QY 1321 CCGGCAACTTCTGTGCAAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 CCGGCAACTTCTGTGCAAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 CCGGCGAGAGCTTCTGTGAGAGAGACCAACCCCGCGCCCAAGAGAGAGAGAGAGAGAG 1440
Db 1381 CCGGCGAGAGCTTCTGTGAGAGAGACCAACCCCGCGCCCAAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 CGGAGCCCTTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1441 CGGAGCCCTTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 AGCCAGTAA 1509
Db 1501 AGCCAGTAA 1509
```

## RESULT 2

US-09-967-464-64

; Sequence 64, Application US/09967464

; Publication No. US20030138453A1

; GENERAL INFORMATION:

; APPLICANT: O'Hagan, Derek

; APPLICANT: Otten, Gillis

; APPLICANT: Donnelly, John J.

; APPLICANT: Polo, John M.

; APPLICANT: Barnett, Susan

; APPLICANT: Singh, Mamohan

; APPLICANT: Ulmer, Jeffrey

; APPLICANT: Dubensky, Jr., Thomas W.

; TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF HETEROLOGOUS NUCLEIC ACIDS

; FILE REFERENCE: P16269, 004

; CURRENT APPLICATION NUMBER: US/09/967,464

; CURRENT FILING DATE: 2002-04-11

; PRIOR APPLICATION NUMBER: 60/236,105

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: 60/315,905

; PRIOR FILING DATE: 2001-08-30

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: Patent version 3.1

; SEQ ID NO 64

; LENGTH: 1509

; TYPE: DNA

; ORGANISM: Human immunodeficiency virus type 1

US-09-967-464-64

Query Match 99.9%; Score 1507.4; DB 10; Length 1509;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1508; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGGCGCCCGGCGAAGACCTTACATGCTGAAGACCTGTGTGGCCACGCCGAG 60
Db 1 ATGGGCGCCCGGCGAAGACCTTACATGCTGAAGACCTGTGTGGCCACGCCGAG 60
QY 61 CTGCGCCCGCGGCAAGACCTTACATGCTGAAGACCTGTGTGGCCACGCCGAG 120
Db 61 CTGCGCCCGCGGCAAGACCTTACATGCTGAAGACCTGTGTGGCCACGCCGAG 120
QY 121 CTGGAAGGGCTTGCCTGTGAACCCCGGCTGTGAGAGACCGCGAGGGCTGTGAAGAGATC 180
Db 121 CTGGAAGGGCTTGCCTGTGAACCCCGGCTGTGAGAGACCGCGAGGGCTGTGAAGAGATC 180
```



QY 181 ATGAAGCAGCTGACGCGCCCTGCAAGACCGGACCGGAGCTGCGCAACCTTTAACAAC 240  
DB 181 ATGAAGCAGCTGACGCGCCCTGCAAGACCGGACCGGAGCTGCGCAACCTTTAACAAC 240  
QY 241 ACCGTGGCACCCTGTACTGCTGTCAGCGCGCATGAGGTCCGCGACCAAGAGAGCC 300  
DB 241 ACCGTGGCACCCTGTACTGCTGTCAGCGCGCATGAGGTCCGCGACCAAGAGAGCC 300  
QY 301 CTGGAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 301 CTGGAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
DB 361 GCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
QY 421 CACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
DB 421 CACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
QY 481 TTCAAGCCCGAGAGTATCCCATTTTACCGCCCTGAGAGAGAGAGAGAGAGAGAG 540  
DB 481 TTCAAGCCCGAGAGTATCCCATTTTACCGCCCTGAGAGAGAGAGAGAGAGAGAG 540  
QY 541 CTGAACAGAGTGTGAACAACCGTGGGCGGCGCAACGAGCGCGCATGAGAGAGAG 600  
DB 541 CTGAACAGAGTGTGAACAACCGTGGGCGGCGCAACGAGCGCGCATGAGAGAGAG 600  
QY 601 ACCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 601 ACCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 661 GCGCGCGGAGAGTCCGCGACCGCGGCGAGCATGCGCGCGCGCAACGAGAGAG 720  
DB 661 GCGCGCGGAGAGTCCGCGACCGCGGCGAGCATGCGCGCGCGCAACGAGAGAG 720  
QY 721 CAGAGAGAGATGCTGCTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
DB 721 CAGAGAGAGATGCTGCTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
QY 781 CCGTGGATCATCTGCGGCTTGAACAAGATCTGCGGATTAACGCGCGCGCATCTG 840  
DB 781 CCGTGGATCATCTGCGGCTTGAACAAGATCTGCGGATTAACGCGCGCGCATCTG 840  
QY 841 GACATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
DB 841 GACATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 901 CTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 901 CTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 CAGAAACGCAACCGCGCATGCAAGACCATCTGCGGCTTGAACAAGAGAGAGAG 1020  
DB 961 CAGAAACGCAACCGCGCATGCAAGACCATCTGCGGCTTGAACAAGAGAGAGAG 1020  
QY 1021 GAGGAGATGATGACCGCTTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
DB 1021 GAGGAGATGATGACCGCTTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1081 GCGGAGAGATGATGACCGCTTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
DB 1081 GCGGAGAGATGATGACCGCTTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1141 GCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
DB 1141 GCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
QY 1201 TGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
DB 1201 TGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

QY 1261 GACTGACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1261 GACTGACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
QY 1321 CCGGCGCACTTCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
DB 1321 CCGGCGCACTTCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
QY 1381 CCGGCGCACTTCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1381 CCGGCGCACTTCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
QY 1441 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
DB 1441 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
QY 1501 AGCCAGTAA 1509  
DB 1501 AGCCAGTAA 1509

RESULT 3  
US-09-967-464-68  
; Sequence 68, Application US/09967464  
; Publication No. US20030138453A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Hagan, Derek  
; APPLICANT: Otten, Gilille  
; APPLICANT: Donnelly, John J.  
; APPLICANT: Polo, John M.  
; APPLICANT: Barnett, Susan  
; APPLICANT: Singh, Mamohan  
; APPLICANT: Ulmer, Jeffrey  
; APPLICANT: Dubensky, Jr., Thomas W.  
; TITLE OF INVENTION: MICROARTICLES FOR DELIVERY OF HETEROLOGOUS NUCLEIC ACIDS  
; FILE REFERENCE: P16269.004  
; CURRENT APPLICATION NUMBER: US/09/967,464  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/236,105  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: 60/315,905  
; PRIOR FILING DATE: 2001-08-30  
; NUMBER OF SEQ ID NOS: 68  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 68  
; LENGTH: 1509  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-967-464-68

Query Match 99.7%; Score 1504.2; DB 10; Length 1509;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 1497; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
DB 1 ATGGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
QY 61 CTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 61 CTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
QY 121 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 121 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
QY 181 ATGAAGCAGCTGACGCGCCCTGCAAGACCGGACCGGAGCTGCGCAACCTTTAACAAC 240  
DB 181 ATGAAGCAGCTGACGCGCCCTGCAAGACCGGACCGGAGCTGCGCAACCTTTAACAAC 240  
QY 241 ACCGTGGCACCCTGTACTGCTGTCAGCGCGCATGAGGTCCGCGACCAAGAGAGCC 300  
DB 241 ACCGTGGCACCCTGTACTGCTGTCAGCGCGCATGAGGTCCGCGACCAAGAGAGCC 300



421 CACCAAGGCGATCAGGCGCGCGCAAGCCCTGAGAGCTGATGATGAGAGAGGCGC 480  
481 TTCAAGCCCGAGGTGATCCCATGTTTACCGCCCTGAGCGAGGCGCCACCCCGCAGAGC 540  
481 TTCAAGCCCGAGGTGATCCCATGTTTACCGCCCTGAGCGAGGCGCCACCCCGCAGAGC 540  
541 CTGAACAGAGTTTGAACAACCGTGGCGGCGCAACAGCGCGGCAATGAGTGTGAAGAGC 600  
541 CTGAACAGAGTTTGAACAACCGTGGCGGCGCAACAGCGCGGCAATGAGTGTGAAGAGC 600  
601 ACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGACACCGCGTGAAGGCGCGCGCTG 660  
601 ACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGACACCGCGTGAAGGCGCGCGCTG 660  
661 GCGCGCGCGAGATGCGGCGAGCCCGCGGCGAGGAGATGCGCGCGCGCGCACAGCACCTG 720  
661 GCGCGCGCGAGATGCGGCGAGCCCGCGGCGAGGAGATGCGCGCGCGCGCACAGCACCTG 720  
721 CAGGAGCAGATGCGCTGAGTGAACAGCAACCCCGCGTCCGCTGGGCGGCAATCTAAG 780  
721 CAGGAGCAGATGCGCTGAGTGAACAGCAACCCCGCGTCCGCTGGGCGGCAATCTAAG 780  
781 CCGTGGATCATCTGCGGCTGGAACAAGATGTCGCGATGTAGAGCCCGTGAAGCATCTG 840  
781 CCGTGGATCATCTGCGGCTGGAACAAGATGTCGCGATGTAGAGCCCGTGAAGCATCTG 840  
841 GACATCCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 900  
841 GACATCCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 900  
901 CTGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 960  
901 CTGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 960  
961 CAGAAAGCGCAACCGCGAGTGAAGACCATCTGCGGCTCTGCGGCGCGCGCGCGCACCTG 1020  
961 CAGAAAGCGCAACCGCGAGTGAAGACCATCTGCGGCTCTGCGGCGCGCGCGCGCACCTG 1020  
1021 GAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
1021 GAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
1081 GCGGAGGAGTGAAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1140  
1081 GCGGAGGAGTGAAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1140  
1141 GCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1200  
1141 GCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1200  
1201 TGCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1260  
1201 TGCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1260  
1261 GACTGCAACGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1320  
1261 GACTGCAACGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1320  
1321 CCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1380  
1321 CCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1380  
1381 CCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1440  
1381 CCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1440  
1441 CCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1500  
1441 CCGCGCGCGCGAGAGGCGCGCAAGAGGCGCTTCCGCGAGTACGTGAGCCGCTTCTTCAAG 1500  
1501 AGCCAGTAA 1509  
1501 AGCCAGTAA 1509

1501 AGCCAGTAA 1509  
RESULT 5  
US-09-899-575-51  
Sequence 51, Application US/09899575  
Publication No. US20030223961A1  
GENERAL INFORMATION:  
APPLICANT: Zur Megele, Jan  
APPLICANT: Barnett, Susan W.  
APPLICANT: Egnelbrecht, Susan  
APPLICANT: van Rensburg, Estrelita Janse  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
FILE REFERENCE: PP01631.102  
CURRENT APPLICATION NUMBER: US/09/899,575  
CURRENT FILING DATE: 2001-07-05  
PRIOR APPLICATION NUMBER: 09/475,704  
PRIOR FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 135  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 51  
LENGTH: 1494  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: HIV Type C  
US-09-899-575-51  
Query Match 86.0%; Score 1297.6; DB 10; Length 1494;  
Best Local Similarity 92.7%; Pred No. 1.3e-271;  
Matches 1402; Conservative 0; Mismatches 89; Indels 21; Gaps 3;  
1 ATGGGCGCGCGCGCGCGAGCATCTGCGCGCGGAGAGCTGAGACAGTGGAGAGATCCGC 60  
1 ATGGGCGCGCGCGCGCGAGCATCTGAGCGCGCGGCAAGCTGAGACAGTGGAGAGATCCGC 60  
61 CTGCGCGCGCGCGCGCGAGAGAGCATCTGAGAGCATCTGAGAGCATCTGAGAGCATCTG 120  
61 CTGCGCGCGCGCGCGCGAGAGAGCATCTGAGAGCATCTGAGAGCATCTGAGAGCATCTG 120  
121 CTGAGGCGCTTGGCGCTGAGACCCCGGCGCTGAGAGAGCGCGGAGGCGCTGAGAGAGT 180  
121 CTGAGGCGCTTGGCGCTGAGACCCCGGCGCTGAGAGAGCGCGGAGGCGCTGAGAGAGT 180  
181 ATGAAGCAGCTGCAACCGCGCGCTGAGAGAGCGCGCAACGAGAGCTGCGAGCTGTACAAC 240  
181 ATGAAGCAGCTGCAACCGCGCGCTGAGAGAGCGCGCAACGAGAGCTGCGAGCTGTACAAC 240  
241 ACCGTGGCGCGCTGTAATGCGCTGCAACCGCGCGCATGAGGTCCGCGACACAGAGAGCC 300  
241 ACCGTGGCGCGCTGTAATGCGCTGCAACCGCGCGCATGAGGTCCGCGACACAGAGAGCC 300  
241 ACCGTGGCGCGCTGTAATGCGCTGCAACCGCGCGCATGAGGTCCGCGACACAGAGAGCC 300  
301 CTGAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
301 CTGAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
361 GCGGAGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
361 GCGGAGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
421 CACGAGGCGATAGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
421 CACGAGGCGATAGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
481 TTCAAGCCCGAGAGTATCCCATGTTTACCGCCCTGAGCGAGGCGCGCGCGCGCGAGAG 540  
481 TTCAAGCCCGAGAGTATCCCATGTTTACCGCCCTGAGCGAGGCGCGCGCGCGCGAGAG 540  
541 CTGAACAGAGTGTGAACAACCGTGGCGGCGCAACAGCGCGGCAATGAGTGTGAAGAGC 600  
541 CTGAACAGAGTGTGAACAACCGTGGCGGCGCAACAGCGCGGCAATGAGTGTGAAGAGC 600

RESULT 6  
US-09-899-575-99  
: Sequence 99, Application US/09899575  
: Publication No. US20030223961A1  
: GENERAL INFORMATION:  
: APPLICANT: Zur Megele, Jan  
: APPLICANT: Barnett, Susan W.

```

1  APPLICANT: Egnelbrecht, Susan
2  APPLICANT: van Rensburg, Estrellita Jame
3  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
4  TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
5  FILE REFERENCE: PP01631.102
6  CURRENT APPLICATION NUMBER: US/09/899,575
7  CURRENT FILING DATE: 2001-07-05
8  PRIOR APPLICATION NUMBER: 09/475,704
9  PRIOR FILING DATE: 1999-12-30
10 NUMBER OF SEQ ID NOS: 135
11 SOFTWARE: PatentIn Ver. 2.0
12 SEQ ID NO 99
13 LENGTH: 1491
14 TYPE: DNA
15 ORGANISM: Artificial Sequence
16 FEATURE:
17 OTHER INFORMATION: Description of Artificial Sequence: Gag_Ty2_C_Znapt
18 US-09-899-575-99
19
20 Query Match      85.8%; Score 1294.8; DB 10; Length 1491;
21 Best Local Similarity 92.7%; Pred. No. 5.2e-271;
22 Matches 1401; Conservative 0; Mismatches 87; Indels 24; Gaps 3
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973

```

Db 721 CAGGAGAGATGCGCTGATGACAGCAACCCCCCATCCCGTGGCGACATCTACAG 780  
Qy 781 CGGTGATATCTCTGGGCTTGAACAAGATCTGCGGATGACGCCCCGTGAGCATCTTG 840  
Db 781 CGCTGGATCTATCTGGGCTTGAACAAGATCTGCGGATGACGCCCCGTGAGCATCTTG 840  
Qy 841 GACATCCGCGAGGCGCCCAAGAGCCCTTCCGGACTAGTGAACCGCTTCTTCAAGACC 900  
Db 841 GACATCCAGAGGCGCCCAAGAGCCCTTCCGGACTAGTGAACCGCTTCTTCAAGACC 900  
Qy 901 CTGCGCGCGGACAGGCGCAACCCAGGACGTGAAGACTGATGACAGGACCTGCTGTG 960  
Db 901 CTGCGCGCGGACAGGCGCAACCCAGGAGGTGAAGACTGATGACAGGACCTGCTGTG 960  
Qy 961 CAGAACCGCAACCCCGGACTGCAAGACATCTGCGGCTCTGAGCCCGCGGCAACCTTG 1020  
Db 961 CAGAACCGCAACCCCGGACTGCAAGACATCTGCGGCTCTGAGCCCGCGGCAACCTTG 1020  
Qy 1021 GAGGAGATGATGACGCGCTGCGGAGGCGTGGGCGGCGCGGCGCAAGGCGCGGCTG 1080  
Db 1021 GAGGAGATGATGACGCGCTGCGGAGGCGTGGGCGGCGCGGCGCAAGGCGCGGCTG 1080  
Qy 1081 GCGGAGCGGATGAGCGGCGCAACAGCTGAAATATCATGATGACAGAGAGCAATTTCAAG 1140  
Db 1081 GCGGAGCGGATGAGCGGCGCAACAGCTGAAATATCATGATGACAGAGAGCAATTTCAAG 1140  
Qy 1141 GCGCGCGCGGCGCAAGCTCAAGTCTTCAACTGCGGCGCAAGAGGCGCAATTCGCAAGAC 1200  
Db 1141 GCGCGCGCGGCGCGCTGAAAGTCTTCAACTGCGGCGCGGCGCAAGAGGCGCAATTCGCAAGAC 1200  
Qy 1201 TGCGCGCGCGCGCGCAAGAGGCGCTGAAAGTGGCGGCAAGAGGCGCAAGATGAAG 1260  
Db 1201 TGCGCGCGCGCGCGCAAGAGGCGCTGAAAGTGGCGGCAAGAGGCGCAAGATGAAG 1260  
Qy 1261 GACTGACCGAGGCGCGCAAGGCTTCTGCGGCAAGATCTGCGGCGCAAGAGGCGCGC 1320  
Db 1261 GACTGACCGAGGCGCGCAAGGCTTCTGCGGCAAGATCTGCGGCGCAAGAGGCGCGC 1320  
Qy 1321 CCGGCGCACTTCTCTCAAGAGCGCGCGCGGCGCGGCGCGCGCGCGCGCGCGCGC 1380  
Db 1321 CCGGCGCACTTCTCTCAAGAGCGCGCGCGGCGCGGCGCGCGCGCGCGCGCGCGC 1380  
Qy 1378 CCGCGCGCGGAGGCTTCCGCTTCAAGAGGCAACCCCGCGCGCGGCGGAGGCGCGCAG 1437  
Db 1381 CCGCGCGCGGAGGCTTCAAGTCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431  
Qy 1438 GACCGCGAGCGCTTACCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1497  
Db 1432 G-----ACCGGAGGCGCGTGAACGAGCTGAAAGAGCTGTGGCGAGGCAAGCCC 1479  
Qy 1498 CTGAGCCAGTAA 1509  
Db 1480 CTGAGCCAGTAA 1491

RESULT 7  
US-09-967-464-63

; Sequence 63, Application US/09967464  
; Publication No. US20030138453A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Hagan, Derek  
; APPLICANT: Otten, Gillis  
; APPLICANT: Donnelly, John J.  
; APPLICANT: Polo, John M.  
; APPLICANT: Barnett, Susan  
; APPLICANT: Singh, Mamohan  
; APPLICANT: Ulmer, Jeffrey  
; APPLICANT: Dubensky, Jr., Thomas W.  
; TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF HETEROLOGOUS NUCLEIC ACIDS  
; FILE REFERENCE: P16269.004  
; CURRENT APPLICATION NUMBER: US/09/967,464  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/236,105

; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: 60/315,905  
; PRIOR FILING DATE: 2001-08-30  
; NUMBER OF SEQ ID NOS: 68  
; SOFTWARE: Patencin version 3.1  
; SEQ ID NO 63  
; LENGTH: 1479  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-967-464-63  
Query Match 84.6%; Score 1276.8; DB 10; Length 1479;  
Best Local Similarity 92.5%; Pred. No. 4,1e-267;  
Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;  
Qy 1 ATGGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60  
Db 1 ATGGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60  
Qy 61 CTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
Db 61 CTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
Qy 121 CTGAGGCGCTTGGCGCTGAAACCCGCGCTGCTGAGAGCCGCGAGGCGCTGCAAGCATC 180  
Db 121 CTGAGGAGGTTGCGCTGAAACCCGCGCTGCTGAGAGCCGCGAGGCGCTGCAAGCATC 180  
Qy 181 ATGAAGAGCTGAGCGCGCGCGCTGCAAGCGGCGCGGAGAGCTGCGGCGCTGTACAC 240  
Db 181 ATCCCGCGAGCTGACCCGCGCGCTGCAAGCGGCGCGGAGAGCTGGAAGAGCTGTACAC 240  
Qy 241 ACCGCGCGCGCGCTGTAACCGCGCTGCAAGCGGCGCGGAGAGCTGCGGCGGCGGCGG 300  
Db 241 ACCGCGCGCGCGCTGTAACCGCGCTGCAAGCGGCGCGGAGAGCTGCGGCGGCGGCGG 300  
Qy 241 ACCGCGCGCGCGCTGTAACCGCGCTGCAAGCGGCGCGGAGAGCTGCGGCGGCGGCGG 300  
Db 241 ACCGCGCGCGCGCTGTAACCGCGCTGCAAGCGGCGCGGAGAGCTGCGGCGGCGGCGG 300  
Qy 301 CTGGAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 301 CTGGAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Qy 361 GCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417  
Db 361 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Qy 418 GTGACACAGGCGCATGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477  
Db 421 GTGACACAGGCGCATGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
Qy 478 GCGTCAAGCGCGGAGGATGATCCCATGTTCAACCGCGCTGAGGAGGCGGCGGCGGCGG 537  
Db 481 GCGTCAAGCGCGGAGGATGATCCCATGTTCAACCGCGCTGAGGAGGCGGCGGCGGCGG 540  
Qy 538 GACCTGAACACGATGTTGAACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597  
Db 541 GACCTGAACACGATGTTGAACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
Qy 598 GACACATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657  
Db 601 GACACATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Qy 658 GTGGCGCGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 717  
Db 661 ATGCGCGCGCGGCGGAGTGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720  
Qy 718 CTGAGAGAGAGATGCTGATGACAGCAACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 777  
Db 721 CTGAGAGAGAGATGCTGATGACAGCAACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780  
Qy 778 AAGCGTGAATCATCTGCGGCTGAAACAAGATCTGCGGAGTGAACAGGCGGCGGCGGCGG 837  
Db 781 AAGCGTGAATCATCTGCGGCTGAAACAAGATCTGCGGAGTGAACAGGCGGCGGCGGCGG 840  
Qy 838 CTGACATCCGCGGAGGCGGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 897

Dd	841	CTGGACATCAACAGGGGCCCCCAAGAGGACCTTCCGCACTACGTTGACCGCTTCTTCAAG	900
Qy	898	ACCTGCGCGCGGAGAGAGCCACCCAGACGTGTAAACTCTGATGACCGAGACCTTGCTG	957
Dd	901	ACCTTGGCGCGCGGAGAGAGACCCAGAGGGTGTAAACCTGTATGACCAACACCTTGCTG	960
Qy	958	GTCAGAAACGCCAACCCCACTGTCAAGAACATCTTGCGCTCTCGGCCCCGGCGCACCC	101
Dd	961	GTCAGAAACGCCAACCCCGACTGTCAAGAACATCTTGCGGCTCTCGGCCCCGGCGCACCC	102
Qy	1018	CTGGAGAGATATATGACCGGCTTGCCCAAGGGCGTGGCGGCCCGGCCACAAAGCCGCGTG	107
Dd	1021	CTGGAGAGATATATGACCGGCTTGCCCAAGGGCGTGGCGGCCCGGCCACAAAGCCGCGTG	108
Qy	1078	CTGGCGGAGCGCATGAGCGAGGCCCAACAGCGTGAACATCATGTATGTCAAGAAAGCACTTC	113
Dd	1081	CTGGCGGAGCGCATGAGCGAGGCCCAACACCC--AGCGTATGATGTCAAGAAAGCACTTC	113
Qy	1138	AAAGGCCCCCGGCGCAACGTCAAGTGTTCAACTGCGGCAAGAGAGGCCACATCGCCAAAG	119
Dd	1138	AAAGGCCCCCGGCGCATGCTCAAGTGTTCAACTGCGGCAAGAGAGGCCACATCGCCCGC	119
Qy	1198	AACGCGCGGCGCCCCCGCAAGAAAGGCTGTGTGAAGTGTGGGCAAGAGAGGCCACCAAGATG	125
Dd	1198	AACGCGCGGCGCCCCCGCAAGAAAGGCTGTGTGAAGTGTGGGCAAGAGAGGCCACCAAGATG	125
Qy	1258	AAGACTGTCAACCGAGCGCGGACCAACTTCTGTGGCAAGATCTGGCCCAAGCAAGAGGC	131
Dd	1258	AAGACTGTCAACCGAGCGCGGACCAACTTCTGTGGCAAGATCTGGCCCAAGCAAGAGGC	131
Qy	1318	CGCCCCGGGCACTTCTCTGTCAAGAACCGGACGAGCCCGCCCCCAACGCTGCCACCGCC	137
Dd	1318	CGCCCCGGGCAACTTCTCTGTGAAGGCGG-----CCCGAGGCCCAACCGCC	135
Qy	1378	CCCCCGCGGAGAGCTTCGCTCTGTGAGAGACCAACCCCGCCCCCAAGAGAGAGCCCAAG	143
Dd	1360	CCCCCGCGGAGAGCTTCGCTCTGTGAGAGACCAACCCCGCGCAAGAGAGAGAGCAAG	141
Qy	1438	GACCGCGAGCCCTTACCGGAGGCGCCTTGAACGCGCTTGGCGCAGCTGTGGCAGCGGCC	149
Dd	1420	G-----ACCGCGAGACCTGTGACAGCGCTTGAAGAGGCTGTGGCAACGACCC	146
Qy	1498	CTGAGCCAGTAA 1509	
Dd	1468	CTGAGCCAGTAA 1479	
RESULT 8			
US-09-899-575-3			
Sequence 3, Application US/09899575			
Publication No. US20030223961A1			
GENERAL INFORMATION:			
APPLICANT: Zur Megele, Jan			
APPLICANT: Barnett, Susan W.			
APPLICANT: Egnelbrecht, Susan			
APPLICANT: van Rensburg, Estrellita Janse			
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C			
FILE REFERENCE: PP01631.102			
CURRENT APPLICATION NUMBER: US/09/899,575			
PRIOR FILING DATE: 2001-07-05			
PRIOR FILING DATE: 1999-12-30			
NUMBER OF SEQ ID NOS: 135			
SOFTWARE: PatentIn Ver. 2.0			
SEQ ID NO 3			
LENGTH: 1479			
TYPE: DNA			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag			
US-09-899-575-3			

Query Match	Similarity	94.6%	Score 1276.8	DB 10	Length 1479
Best Local	Similarity 92.5%	Pred. No. 4,1e-267			
Matches 1399	Conservative 0	Mismatches 77	Indels 36	Gaps 4	
QY	1	ATGGCGCCCGCGCCAGCATCTCGCGCGCGAGAGCTTGACAAAGTGAGAGAAATCCGC	60		
Db	1	ATGGCGCGCCCGCGCCAGCATCTCGCGCGCGCGAGAGCTTGACCGCTTGAGACCGATCCGC	60		
QY	61	CTGGCGCCCGCGCGCGAGAAAGCACTAATGCTGAAGCACTGATGTGGGCCAGCCGAG	120		
Db	61	CTGGCGCCCGCGCGCGAGAAAGCTCACTAATGAAGCACTGATGTGGGCCAGCCGAG	120		
QY	121	CTGAGAGGCTTCGCCCTGATACCCCGCGCGCTGTGAGACCGCGAGAGGCTGCAACATC	180		
Db	121	CTGAGAGAGTTTCGCCCTGATACCCCGCGCGCTGTGAGACCAAGAGAGGCTGCAACATC	180		
QY	181	ATGAAGCAGCTGACAGCCCGCGCTGACACCGGACCGAGAGCTGCGACGCTTGAAC	240		
Db	181	ATCCCGCAGCTGACACCCCGCGCTGACACCGGACCGAGAGCTGAAGAGCTTGAAC	240		
QY	241	ACCTGCGCACCTCTGTACTGCGTGAACGCGCGGACATCGAGGTCGCGGACCAAGAGCC	300		
Db	241	ACCTGCGCACCTCTGTACTGCGTGAACGCGCGGACATCGAGGTCGCGGACCAAGAGCC	300		
QY	301	CTGACAAAGATCGAGAGAGCAGAACAAAGTCCGACGAAGAACCCAGAGCCAGAGAG	360		
Db	301	CTGACAAAGATCGAGAGAGCAGAACAAAGTCCGACGAAGAACCCAGAGCCAGAGAG	360		
QY	361	GCGGCA---GGCAAGGTGAGCGCAGAACAAACCCATCTGTCGAACCTGCGAGGCAATG	417		
Db	361	GCGGCAAGGAGGAGGTGAGCGCAGAACAAACCCATCTGTCGAACCTGCGAGGCAATG	420		
QY	418	GTGACCAAGGCCATCAGGCCCGCGGACCTCTGACGCTGTGAGAGGTATCGAGAGAG	477		
Db	421	GTGACCAAGGCCATCAGGCCCGCGGACCTCTGACGCTGTGAGAGGTATCGAGAGAG	480		
QY	478	GCTTCAAGCCCCGAGGTATCTCCATGTTCAACGCGCTGAGCGAGGGGCCACCCCGAG	537		
Db	481	GCTTCAAGCCCCGAGGTATCTCCATGTTCAACGCGCTGAGCGAGGGGCCACCCCGAG	540		
QY	538	GACCTGAACAGATGTTGAACACCGTGGGCGGCGGACAGAGCGCGCATGAGATCTGAAG	597		
Db	541	GACCTGAACAGATGTTGAACACCGTGGGCGGCGGACAGAGCGCGCATGAGATCTGAAG	600		
QY	598	GACACCATCAACGAGAGAGCGCGGAGTGGACCGCTGACACCCGTGACAGCGGCGCC	657		
Db	601	GACACCATCAACGAGAGAGCGCGGAGTGGACCGCTGACACCCGTGACAGCGGCGCC	660		
QY	658	GTTGGCCCCCGGCGAGTGGCGAACCCCGGCGGAGGCAATCGCGCGCGCACACAGACC	717		
Db	661	ATTGGCCCCCGGCGAGTGGCGAACCCCGGCGGAGGCAATCGCGCGCACACAGACC	720		
QY	718	CTGACAGAGCAGATCGCTGTGATGACAGAACCCCGCGTGGCCCTGTGGCGCATCTAC	777		
Db	721	CTGACAGAGCAGATCGCTGTGATGACAGAACCCCGCATCCCGTGTGGCGCATCTAC	780		
QY	778	AAAGCGTGTGATCAATCTGGGCTTGAACAAAGTGTGGATGTACAGCCCGTGAGCATC	837		
Db	781	AAAGCGTGTGATCAATCTGGGCTTGAACAAAGTGTGGATGTACAGCCCGTGAGCATC	840		
QY	838	CTGACATCCGCGCAGAGGCGCCCAAGAGCCCTTCGCGCATACGAGAGCCGCTTTCAAG	897		
Db	841	CTGACATCCGCGCAGAGGCGCCCAAGAGCCCTTCGCGCATACGAGAGCCGCTTTCAAG	900		
QY	898	ACCTGCGCGCGAGAGAGCCACCGAGAGCTGAAGAACTGATGACCGAGACCCCTGCTG	957		
Db	901	ACCTGCGCGCGAGAGAGCCACCGAGAGTGAAGAACTGATGACCGAGACCCCTGCTG	960		
QY	958	GTGAGAAAGCGCAACCCCGCATGGAACAATCTGCGCGCTCTCGGCGCCCGGCGCAC	1017		
Db	961	GTGAGAAAGCGCAACCCCGCATGGAACAATCTGCGCGCTCTCGGCGCCCGGCGCAC	1020		

QY 1018 CTGAGAGATGATGACCGCTGCGAGGGGCTGGGGGCGCCCGGCGCAAGAGCCGCGG 1077  
DB 1021 CTGAGAGATGATGACCGCTGCGAGGGGCTGGGGGCGCCCGGCGCAAGAGCCGCGG 1080  
QY 1078 CTGGCGAGGCGATGAGCGAGGCGCAACAGCTGGAACATGATGATGAGAGCACTTC 1137  
DB 1081 CTGGCGAGGCGATGAGCGAGGCGCAACAGCTGGAACATGATGATGAGAGCACTTC 1137  
QY 1138 AGGGGCGCGGCGCAAGCTCAAGTCTTCAACTGCGGCAAGAGGGCGCACTCGCCAG 1197  
DB 1138 AGGGGCGCGGCGCAAGCTCAAGTCTTCAACTGCGGCAAGAGGGCGCACTCGCCAG 1197  
QY 1198 AACTGCGGCGCGCGCGCAAGAGGGCTCTGGAAGTGGGCGCAAGAGGGCGCAAGATG 1257  
DB 1198 AACTGCGGCGCGCGCGCAAGAGGGCTCTGGAAGTGGGCGCAAGAGGGCGCAAGATG 1257  
QY 1258 AAGGACTGCAAGGAGCGGCGCAACTTCTGGGCAAGATGCGCCAGCGCAAGAGGCG 1317  
DB 1258 AAGGACTGCAAGGAGCGGCGCAACTTCTGGGCAAGATGCGCCAGCGCAAGAGGCG 1317  
QY 1318 CGCGCGGCGCACTTCTGCAAGACCGGAGAGGCGCGCGCGCGCGCGCGCGCGCGCG 1377  
DB 1318 CGCGCGGCGCACTTCTGCAAGACCGGAGAGGCGCGCGCGCGCGCGCGCGCGCGCG 1359  
QY 1378 CGCGCGGCGGAGGCTTCTGCGAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437  
DB 1360 CGCGCGGCGGAGGCTTCTGCGAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1419  
QY 1438 GACCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1497  
DB 1420 G-----ACCGGAGAGCGCTGACCGAGCTGAAGAGCGCTGTCGCGCAAGCGCGCG 1467  
QY 1498 CTGAGCGAGTAA 1509  
DB 1468 CTGAGCGAGTAA 1479

RESULT 9  
US-10-190-435-18  
; Sequence 18, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 3162  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagRtmuc\_C  
US-10-190-435-18

Query Match 84.5%; Score 1275.2; DB 15; Length 3162;  
Best Local Similarity 92.5%; Pred. No. 8.9e-267;  
Matches 1398; Conservative 0; Mismatches 78; Indels 36; Gaps 4;  
QY 1 AAGGGGCGCGGCGCGCAAGCTTCTGGCGGCGGCGAGAACCTGCAAGAGGAGAAATCCGC 60  
DB 7 AAGGGGCGCGGCGCGCAAGCTTCTGGCGGCGGCGAGAACCTGCAAGAGGAGAAATCCGC 66  
QY 61 CTGGCGCGCGGCGCGCAAGAGCACTACATGCTGAAGCACTGGTGTGGCGCGAGCGCGAG 120  
DB 67 CTGGCGCGCGGCGCGCAAGAGCACTACATGCTGAAGCACTGGTGTGGCGCGAGCGCGAG 126

QY 121 CTGAGAGGCTTGGCCCTGGAACCCCGGCGCTGAGAGACCGCGGCGCTGCAAGCAATC 180  
DB 127 CTGAGAGAGTGGCCCTGGAACCCCGGCGCTGAGAGACCGCGGCGCTGCAAGCAATC 186  
QY 181 ATGAAGCAGCTGACAGCCCGCTGCGAGACCGGCGCAAGAGAGAGCTGCGACCTGTACAC 240  
DB 187 ATCCGCGAGCTGACACCCCGCTGCGAGACCGGCGCAAGAGAGAGCTGAGAGAGCTGTAC 246  
QY 241 ACCGTGGCGCGCTGTACTGCGGTGCAACCGCGGCGATGAGGTCCGCGACCAAGAGGCC 300  
DB 247 ACCGTGGCGCGCTGTACTGCGGTGCAAGAGAGAGAGAGAGGTGCGCGACCAAGAGGCC 306  
QY 301 CTGGAACAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGGAACAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCGGAC---GGCAAGGTGAGCGAGAACTACCTGCTGTCAGAACTTGCAGAGGCGCAGATG 417  
DB 367 GCGGACAGAGGCGAGAGGTGAGCGAGAACTACCTGCTGTCAGAACTTGCAGAGGCGCAG 426  
QY 418 GTGACACAGGCGCATGAGCCCGCGCAACCTTGAACGCTGTGGTGAAGGTATGAGAGAG 477  
DB 427 GTGACACAGGCGCATGAGCCCGCGCAACCTTGAACGCTGTGGTGAAGGTATGAGAGAG 486  
QY 478 GCGTTGAGCGCGGAGGTATCCCGATGTTCAACCGCTGTAGAGGAGGCGCGCGCGCG 537  
DB 487 GCGTTGAGCGCGGAGGTATCCCGATGTTCAACCGCTGTAGAGGAGGCGCGCGCGCGCG 546  
QY 538 GACCTGAACAGATGTTGAACAACCGTGGGCGGCGCAACAGCGCGCGCATGAGAGTCTGAG 597  
DB 547 GACCTGAACAGATGTTGAACAACCGTGGGCGGCGCAACAGCGCGCGCATGAGAGTCTGAG 606  
QY 598 GACACATCAAGAGAGAGGCGCGGAGTGGAGCGCGCTGACACCGCGTGAAGCGCGCGCG 657  
DB 607 GACACATCAAGAGAGAGGCGCGGAGTGGAGCGCGCTGACACCGCGTGAAGCGCGCGCG 666  
QY 658 GTGGCGCGGCGCAAGTGGCGGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 717  
DB 667 ATCGCGCGGCGCAAGTGGCGGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 726  
QY 718 CTGAGAGAGCAGATGCGCTGATGAGCAAGCAACCGCGCGTCCGCTGGGCGGCGCATCTAC 777  
DB 727 CTGAGAGAGCAGATGCGCTGATGAGCAAGCAACCGCGCGTCCGCTGGGCGGCGCATCTAC 786  
QY 778 AAGCGGTGATCATCTGGGCGTGAACAAGATGTCGGATGTACAGCCCGTGAAGCATC 837  
DB 787 AAGCGGTGATCATCTGGGCGTGAACAAGATGTCGGATGTACAGCCCGTGAAGCATC 846  
QY 838 CTGACATTCGCGAGGCGCGCAAGAGCGCTTCCGCGACTACGTGAGCGGCTTTCAAG 897  
DB 847 CTGACATTCGCGAGGCGCGCAAGAGCGCTTCCGCGACTACGTGAGCGGCTTTCAAG 906  
QY 898 ACCCTGGCGCGGAGAGAGGCGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957  
DB 907 ACCCTGGCGCGGAGAGAGGCGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966  
QY 958 GTGAGAAAGCGCAACCGCGCTGCAAGACCATCTGCGGCTTCTGCGCGCGCGCGCGCAC 1017  
DB 967 GTGAGAAAGCGCAACCGCGCTGCAAGACCATCTGCGGCTTCTGCGCGCGCGCGCGCAC 1026  
QY 1018 CTGAGAGAGATGATGAGCGGCTGCGAGGCGGCGGCGCGCGCGCGCGCGCGCGCG 1077  
DB 1027 CTGAGAGAGATGATGAGCGGCTGCGAGGCGGCGGCGCGCGCGCGCGCGCGCGCG 1086  
QY 1078 CTGGCGAGGCGATGAGCGGCGCAACAGCGTGAACATCATGATGAGAGAGCACTTC 1137  
DB 1087 CTGGCGAGGCGATGAGCGGCGCAACAGCGTGAACATCATGATGAGAGAGCACTTC 1143  
QY 1138 AAGGGCGCGGCGGCAAGCTCAAGTCTTCAACTGCGGCAAGAGAGGCGCAATCGCCAG 1197  
DB 1144 AAGGGCGCGGCGGCAAGCTCAAGTCTTCAACTGCGGCAAGAGAGGCGCAATCGCCAG 1203



QY 1198 AACTGCGCGCCCGCCGCAAGAGGCTGTGAGAGTGCGGCAAGAGGCGCACGAGATG 1257  
DB 1204 AACTGCGCGCGCCCGCCGCAAGAGGCTGTGAGAGTGCGGCAAGAGGCGCACGAGATG 1263  
QY 1258 AAGAGCTGCAACCGAGCGCCAGGCGCACTTCTGTGGGCAAGATCTGGCCCAAGGAGC 1317  
DB 1264 AAGAGCTGCAACCGAGCGCCAGGCGCACTTCTGTGGGCAAGATCTGGCCCAAGGAGC 1323  
QY 1318 CGCCCGCGCACTTCTGTGAGAGAGCGGAGAGCGCCCGCCCGCACCGTGTCCCAACCGC 1377  
DB 1324 CGCCCGCGCACTTCTGTGAGAGCGG-----CGCCGAGCGCCCGCC 1365  
QY 1378 CCCCCCGCGAGAGCTTCTGTGAGAGAGCAACCGCCCGCCCGCAAGAGAGGCGCAAG 1437  
DB 1366 CCCCCCGCGAGAGCTTCTGTGAGAGAGCAACCGCCCGCCCGCAAGAGAGGAGCAG 1425  
QY 1438 GACCGGAGCGCTTACCGGAGCGCCCTGACCGCCCTGCGAGCGCTGTGGCAGCGGCGCC 1497  
DB 1426 G-----ACCGGAGAGCGCTGAGCAAGCGCTGTAAGAGCGCTGTGGCAAGCGCC 1473  
QY 1498 CTGAGCGCAGTAA 1509  
DB 1474 CTGAGCGCAGAAA 1485

RESULT 10  
US-10-190-435-16  
Sequence 16, Application US/10190435  
Publication No. US20030143248A1  
GENERAL INFORMATION:  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: BARNETT, Susan W.  
APPLICANT: LIAN, Ying  
APPLICANT: ENGELBRECHT, Susan  
APPLICANT: VAN KENSBURG, Estrellita J.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
FILE REFERENCE: P18133.003 / 2302-18133  
CURRENT APPLICATION NUMBER: US/10/190,435  
NUMBER OF SEQ ID NOS: 319  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 16  
LENGTH: 3462  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURES:  
OTHER INFORMATION: Description of Artificial Sequence: GagProteinArtmut\_C  
US-10-190-435-16

Query Match 84.5%; Score 1275.2; DB 15; Length 3462;  
Best Local Similarity 92.5%; Pred. No. 8.8e-267;  
Matches 1398; Conservative 0; Mismatches 78; Indels 36; Gaps 4;  
QY 1 ATGGGCGCGCGCGCGCAGATCTGTGCGCGGAGAGCTGAGCAAGTGGGAGAAATCCGC 60  
DB 7 ATGGGCGCGCGCGCGCAGATCTGTGCGCGGCGGCAAGCTGAGCGCTGGGAGGCGATCCGC 66  
QY 61 CTGCGCGCGCGCGGCAAGAGCACTAATCTGGAAGCACTGTGTGTGGCGCCAGCGCGAG 120  
DB 67 CTGCGCGCGCGGCAAGAGCACTAATCTGGAAGCACTGTGTGTGGCGCCAGCGCGAG 126  
QY 121 CTGAGGCGCTTGTGCTGAACCGCGCGCTGTGTGAGAGCGCGGAGGCGTGAAGCAATC 180  
DB 127 CTGGAAGAGTGTGCTTGAACCGCGCGCTGTGTGAGAGCGAGGAGGCTGCAAGCAATC 186  
QY 181 ATGAAGCACTGCAACCGCGCGCTGTGAGACCGGCAACGAGAGCTGCGAGCGCTGTACAC 240  
DB 187 ATCGCGCAGCTGCAACCGCGCGCTGTGAGACCGGCAACGAGAGCTGAAGAGCTGTTCAC 246  
QY 241 ACCGTGCGCACTCTTACTGTGTGTGACCGCGCGCATGAGGTCCGCGACCAAGAGGCGC 300  
DB 247 ACCGTGCGCACTCTTACTGTGTGTGACGAGAGATGAGGTCCGCGACCAAGAGGCGC 306

QY 301 CTGGAAGAGTGTGAGAGAGGAGCAAAAGTTCAGCAAGAGAGCCAGAGGCGCAAGAG 360  
DB 307 CTGGAAGAGTGTGAGAGAGGAGCAAAAGTTCAGCAAGAGAGATTCAGAGGCGCGAGCC 366  
QY 361 GCCGAC---GGCAAGTGTGAGGAGCACTAACCAGTGTGAGAGAACTGTGAGGCGCAGATG 417  
DB 367 GCCGACAAAGGCGAAGTGTGAGGAGCACTAACCAGTGTGAGAGAACTGTGAGGCGCAGATG 426  
QY 418 GTGACCAAGGCGCATGAGCGCCCGCAACCTGAACGCTGTGGTGAAGTGTGAGAGAGAG 477  
DB 427 GTGACCAAGGCGCATGAGCGCCCGCAACCTGAACGCTGTGGTGAAGTGTGAGAGAGAG 486  
QY 478 GCTTCAAGCGCGGAGTGTATCCCATTTTCAACCGGCTGTGAGAGAGGCGCGCAACCGCCAG 537  
DB 487 GCTTCAAGCGCGGAGTGTATCCCATTTTCAACCGGCTGTGAGAGAGGCGCGCAACCGCCAG 546  
QY 538 GACCTGAACAGATTTGAACACCGTGGGCGGCGCACAGGCGCGCATGACAGATGCTGAG 597  
DB 547 GACCTGAACAGATTTGAACACCGTGGGCGGCGCACAGGCGCGCATGACAGATGCTGAG 606  
QY 598 GACACCATCAACGAGAGAGCGCGGAGTGGAGACCGCTGCAACCGGTGCAAGGCGCGCC 657  
DB 607 GACACCATCAACGAGAGAGCGCGGAGTGGAGACCGCGGTGCAACCGGTGCAAGGCGCGCC 666  
QY 658 GTGGCGCGCGCGCAGATGCGCGGAGCGCGCGCGGCGGAGCATGCGCGCGCGCAACGAC 717  
DB 667 ATGCGCGCGCGCGCAGATGCGCGGAGCGCGCGCGGCGGAGCATGCGCGCGCGCAACGAC 726  
QY 718 CTGCAAGAGAGATGCGCTGTGATGACGAGCAACCGCGCGCGGTGGGAGCATCTAC 777  
DB 727 CTGCAAGAGAGATGCGCTGTGATGACGAGCAACCGCGCGCGGTGGGAGCATCTAC 786  
QY 778 AAGCGGTGATCTCTGTGGCTTGAACAAGATGTGTGGATGTACAGCGCGGTGAGATC 837  
DB 787 AAGCGGTGATCTCTGTGGCTTGAACAAGATGTGTGGATGTACAGCGCGGTGAGATC 846  
QY 838 CTGGAATCTCGCGCAGAGGCGCGCAAGAGAGCGCTTCCGCGCATCTGTGAGAGCGCTTCAAG 897  
DB 847 CTGGAATCTCGCGCAGAGGCGCGCAAGAGAGCGCTTCCGCGCATCTGTGAGAGCGCTTCAAG 906  
QY 898 ACCCTGCGCGCGCGAGAGGCGCACCGAGAGCGTGAAGACTGATGACCGAGACCGCTGTG 957  
DB 907 ACCCTGCGCGCGCGAGAGAGAGCGCACCGAGAGGTAAGACTGATGACCGAGACCGCTGTG 966  
QY 958 GTGCAAGAGCGCAACCGCGACTGTGAGAGCAATCTGTGCGGCTGTGAGCGCGCGGCGCAC 1017  
DB 967 GTGCAAGAGCGCAACCGCGACTGTGAGAGCAATCTGTGCGGCTGTGAGCGCGCGGCGCAC 1026  
QY 1018 CTGGAAGAGATGATGACCGCTGTGCAAGGCGGTGGCGCGCGCGCGCGCAAGCGCGCGGTG 1077  
DB 1027 CTGGAAGAGATGATGACCGCTGTGCAAGGCGGTGGCGCGCGCGCGCGCGCAAGCGCGCGGTG 1086  
QY 1078 CTGCGCGAGGCGATGAGCGCGCAAGCGGTGAACATCTGATGAGCAAGAGAGCACTTC 1137  
DB 1087 CTGCGCGAGGCGATGAGCGCGCAAGCGGTGAACATCTGATGAGCAAGAGAGCACTTC 1143  
QY 1138 AAGGCGCGCGCGCAACGTCAAGTCTTCACTGTGCGCAAGAGGCGCAATGCGCAAG 1197  
DB 1144 AAGGCGCGCGCGCAACGTCAAGTCTTCACTGTGCGCAAGAGGCGCAATGCGCGCG 1203  
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGTGCGCAAGAGGCGCGCAACAGATG 1257  
DB 1204 AACTGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGTGCGCAAGAGGCGCGCAACAGATG 1263  
QY 1258 AAGAGCTGCAACCGAGCGCCAGGCGCACTTCTGTGGGCAAGATCTGGCCCAAGGAGC 1317  
DB 1264 AAGAGCTGCAACCGAGCGCCAGGCGCACTTCTGTGGGCAAGATCTGGCCCAAGGAGC 1323  
QY 1318 CGCCCGCGCACTTCTGTGAGAGAGCGGAGAGCGCCCGCCCGCACCGTGTCCCAACCGC 1377  
DB 1324 CGCCCGCGCACTTCTGTGAGAGAGCGGAGAGCGCCCGCCCGCACCGTGTCCCAACCGC 1386





RESULT 12  
US-10-190-305A-14  
; Sequence 14, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190,305A  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 4419  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: GagRTmutatRevNef\_C  
US-10-190-305A-14

Query Match 84.5%; Score 1275.2; DB 16; Length 4419;  
Best Local Similarity 92.5%; Pred. No. 8.7e-267; Mismatches 78; Indels 36; Gaps 4;  
Matches 1398; Conservative 0;

QY 1 ATGGGCGCCCGCGCCAGCATCTCTGCGCGGAGAGAGCTGAGCAAGTGAGAGAGATCCGC 60  
DB 7 ATGGGCGCCCGCGCCAGCATCTCTGCGCGGAGAGAGCTGAGCAAGTGAGAGAGATCCGC 66  
QY 61 CTGGCGCCCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 67 CTGGCGCCCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126  
QY 121 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 127 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186  
QY 181 ATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 187 ATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
QY 241 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
DB 247 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
QY 301 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCGGAC---GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 417  
DB 367 GCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426  
QY 418 GTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 477  
DB 427 GTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486  
QY 478 GCGTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 537  
DB 487 GCGTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546  
QY 538 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597  
DB 547 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606  
QY 598 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657  
DB 607 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666

QY 658 GTGGCGCCCGCGCGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717  
DB 667 ATGCGCCCGCGCGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
QY 718 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 777  
DB 727 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786  
QY 778 AAGCGGTGATCATCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 837  
DB 787 AAGCGGTGATCATCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 846  
QY 838 CTGAGATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897  
DB 847 CTGAGATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906  
QY 898 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957  
DB 907 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966  
QY 958 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017  
DB 967 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026  
QY 1018 CTGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077  
DB 1027 CTGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086  
QY 1078 CTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137  
DB 1087 CTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1143  
QY 1138 AAGGCGCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197  
DB 1144 AAGGCGCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1203  
QY 1198 AACTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
DB 1204 AACTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1263  
QY 1258 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317  
DB 1264 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1323  
QY 1318 CGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377  
DB 1324 CGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1385  
QY 1378 CCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
DB 1386 CCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1445  
QY 1438 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
DB 1446 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1505  
QY 1498 CTGAGCGAGTAA 1509  
DB 1474 CTGAGCGAGAAA 1485

RESULT 13  
US-10-190-435-35  
; Sequence 35, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF





QY 61 CTGCGCCCCGGCGCAAGAGCATATGCTGAAGCACCTGGTGTGGGCGAGCGCGAG 120  
DB 3197 CTGCGCCCCGGCGCAAGAGCATATGCTGAAGCACCTGGTGTGGGCGAGCGCGAG 3256  
QY 121 CTGGAAGGCTTGGCCCTGGAACCCCGGCTGTGAGAACCGCGAGGGCTGCAAGCATATC 180  
DB 3257 CTGGAAGAGTTGGCCCTGGAACCCCGGCTGTGAGAACCGAGAGGGCTGCAAGCATATC 3316  
QY 181 ATGAAGCAGCTGCAAGCCCGCTGTGCAAGCCGGCAAGAGAGTGGCGACCTGTACAAAC 240  
DB 3317 ATCCCGCAGCTGCAAGCCCGCTGTGCAAGCCGGCAAGAGAGTGGCGACCTGTACAAAC 3376  
QY 241 ACCGTGGACCCCTGACTGCGGTGCAAGCCGGCATGAGGTCCGGGCAACCAAGAGAGGC 300  
DB 3377 ACCGTGGACCCCTGACTGCGGTGCAAGAGAGTGAAGTCCGGGCAACCAAGAGAGGC 3436  
QY 301 CTGGAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 3437 CTGGAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3496  
QY 361 GCGGAC---GGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 417  
DB 3497 GCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3556  
QY 418 GTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 477  
DB 3557 GTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3616  
QY 478 GCGTTCAAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 537  
DB 3617 GCGTTCAAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3676  
QY 538 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597  
DB 3677 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3736  
QY 598 GACACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657  
DB 3737 GACACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3796  
QY 658 GTGGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717  
DB 3797 ATGCGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3856  
QY 718 CTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 777  
DB 3857 CTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3916  
QY 778 AAGCGGTGAGATCATCTGGGCGCTGAACAGAGATGATGCGGATGTACAGCCCGGTGAGCATC 837  
DB 3917 AAGCGGTGAGATCATCTGGGCGCTGAACAGAGATGATGCGGATGTACAGCCCGGTGAGCATC 3976  
QY 838 CTGCAATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897  
DB 3977 CTGCAATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4036  
QY 898 ACCCTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957  
DB 4037 ACCCTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4096  
QY 958 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017  
DB 4097 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4156  
QY 1018 CTGGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1077  
DB 4157 CTGGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4216  
QY 1078 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137  
DB 4217 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4273  
QY 1138 AAGGCCCCCGGCGCAAGTCAAGTGTCTCAACTGGGCGAAGAGAGAGAGAGAGAGAGAGAG 1197

DB 4274 AAGGCCCCCGGCGCATGCTCAAGTGTCTCAACTGGGCGAAGAGAGAGAGAGAGAGAGAGAG 4333  
QY 1198 AACTCCGCGCGCCCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
DB 4334 AACTCCGCGCGCCCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4393  
QY 1258 AAGGACTGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317  
DB 4394 AAGGACTGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4453  
QY 1318 CGCCCGGCACTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377  
DB 4454 CGCCCGGCACTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4495  
QY 1378 CCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
DB 4496 CCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4555  
QY 1438 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
DB 4556 G-----ACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4603  
QY 1498 CTGAGCCAGTAA 1509  
DB 4604 CTGAGCCAGTAA 4615

Search completed: March 13, 2005, 04:29:43  
Job time : 923.106 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:39:57 ; Search time 178.272 Seconds

(without alignments)  
5584.002 Million cell updates/sec

Title: US-09-475-704A-4

Perfect score: 2840

Sequence: 1 atggcgccgcgcgcagcat.....gcgcgccttgaccagtraa 1509

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1396920 seqs, 329844858 residues  
Total number of hits satisfying chosen parameters: 2793840

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlp  
-O=/cgn2\_1/USPTO\_epool\_p/US09475704/runatc\_10032005\_140224\_14937/app\_query.fasta\_1.3342  
-DB=Published Applications\_AA -OPMT=faetan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09475704@cgn2\_1.1.466@runatc\_10032005\_140224\_14937  
-NCPU=6 -ICPU=3 -NO\_MMWP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pdb.\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pdb.\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pdb.\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pdb.\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pdb.\*  
6: /cgn2\_6/ptodata/1/pubppaa/PCTUS\_PUBCOMB.pdb.\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pdb.\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pdb.\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09A\_PUBCOMB.pdb.\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09B\_PUBCOMB.pdb.\*  
11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pdb.\*  
12: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pdb.\*  
13: /cgn2\_6/ptodata/1/pubppaa/US10A\_PUBCOMB.pdb.\*  
14: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pdb.\*  
15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pdb.\*  
16: /cgn2\_6/ptodata/1/pubppaa/US10D\_PUBCOMB.pdb.\*  
17: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pdb.\*  
18: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB.pdb.\*  
19: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pdb.\*  
20: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pdb.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result Query

## SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	2661	93.7	502	10	US-09-899-575-22	Sequence 22, Appl
2	2464	86.8	492	16	US-10-332-413-2	Sequence 2, Appl1
3	2457	86.5	492	9	US-09-991-258-5	Sequence 5, Appl1
4	2457	86.5	631	16	US-10-332-413-8	Sequence 8, Appl1
5	2440	85.9	492	14	US-10-339-217-109	Sequence 109, App
6	2430	85.6	492	14	US-10-339-217-110	Sequence 110, App
7	2421.5	85.3	491	14	US-10-339-217-140	Sequence 140, App
8	2413.5	85.0	491	14	US-10-339-217-143	Sequence 143, App
9	2387	84.0	508	14	US-10-339-217-105	Sequence 105, App
10	2385.5	84.0	491	14	US-10-339-217-104	Sequence 104, App
11	2375.5	83.6	497	14	US-10-339-217-107	Sequence 107, App
12	2353.5	82.9	487	14	US-10-339-217-106	Sequence 106, App
13	2348	82.7	496	14	US-10-339-217-108	Sequence 108, App
14	2343	82.5	492	10	US-09-899-575-17	Sequence 17, Appl
15	2258	79.5	500	14	US-10-339-217-18	Sequence 18, Appl
16	2247.5	79.1	494	14	US-10-339-217-126	Sequence 126, Appl
17	2246	79.1	500	14	US-10-223-172A-42	Sequence 42, Appl
18	2246	79.1	500	14	US-10-224-999A-3483	Sequence 3483, Ap
19	2238	78.8	1350	10	US-09-952-060-35	Sequence 35, Appl
20	2238	78.8	1350	16	US-10-380-641-35	Sequence 35, Appl
21	2232	78.6	557	14	US-10-339-217-144	Sequence 144, App
22	2226	78.4	500	16	US-10-335-468-8	Sequence 8, Appl1
23	2225.5	78.4	494	14	US-10-339-217-100	Sequence 100, App
24	2223	78.3	500	14	US-10-097-534-32	Sequence 32, Appl
25	2223	78.3	500	14	US-10-059-271-85	Sequence 85, Appl
26	2223	78.3	500	14	US-10-102-622-4	Sequence 4, Appl1
27	2223	78.3	500	16	US-10-325-468-34	Sequence 34, Appl1
28	2223	78.3	500	17	US-10-844-658-6	Sequence 6, Appl1
29	2221	78.2	500	9	US-09-968-355-26	Sequence 26, Appl
30	2221	78.2	583	9	US-09-968-355-17	Sequence 17, Appl
31	2220.5	78.2	499	17	US-10-844-658-2	Sequence 2, Appl1
32	2220	78.2	498	16	US-10-325-468-22	Sequence 22, Appl
33	2216.5	78.0	494	14	US-10-339-217-99	Sequence 99, Appl
34	2212.5	77.9	501	14	US-10-339-217-136	Sequence 136, App
35	2212.5	77.9	503	14	US-10-339-217-127	Sequence 127, App
36	2212	77.9	500	14	US-10-339-217-122	Sequence 122, App
37	2212	77.9	500	16	US-10-325-468-17	Sequence 17, Appl
38	2209.5	77.8	494	14	US-10-339-217-101	Sequence 101, App
39	2209.5	77.8	499	15	US-10-296-734-1	Sequence 1, Appl1
40	2205	77.7	496	14	US-10-339-217-32	Sequence 32, Appl
41	2205	77.6	498	16	US-10-325-468-27	Sequence 27, Appl
42	2199	77.4	500	14	US-10-339-217-121	Sequence 121, App
43	2197	77.4	512	16	US-10-093-953A-25	Sequence 25, Appl
44	2197	77.4	512	16	US-10-093-953A-30	Sequence 30, Appl
45	2193	77.2	493	14	US-10-339-217-111	Sequence 111, App

## ALIGNMENTS

RESULT 1  
US-09-899-575-22  
; Sequence 22, Application US/09899575  
; Publication No. US20030223961A1  
GENERAL INFORMATION:  
; APPLICANT: Zur Megecke, Jan  
; APPLICANT: Barnette, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Estrelita Janse  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P01631.102  
; CURRENT APPLICATION NUMBER: US/09/899, 575  
; PRIOR FILING DATE: 2001-07-05  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Human Immunodeficiency Virus







US-09-991-258-5

## Alignment Scores:

Pred. No.:	2,298-118	Length:	492
Score:	2457.00	Matches:	465
Percent Similarity:	95.22%	Conservative:	13
Best Local Similarity:	92.63%	Mismatches:	14
Query Match:	86.51%	Indels:	10
DB:	9	Gaps:	2

US-09-475-704A-4 (1-1509) x US-09-991-258-5 (1-492)

```

QY 1 ATGGGCGCCGCGCGCATCTCTGGCGGCGAGACCTGTGAACAAGTGGAGAAATCCGC 60
   |||
Db 1 MetAlaAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGlyLysIleArg 20
QY 61 CTGGCGCCGCGGCGCAAGAGCACTACATGTGTGAACACCTGTGTGGCCAGCCGCGAG 120
   |||
Db 21 LeuArgProGlyGlyLysLysIleTrpMetLeuLysIleValTrpAlaSerArgGly 40
QY 121 CTGGAGGGCTTGGCCCTGAACCCCGGCTGTGTGAGACCGCGAGGGCTGTCAAGATC 180
   |||
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlyGlyCysLysGlnIle 60
QY 181 ATGAAGCAGCTGACGCGCCCTGTGACAGCGGACCGAGAGCTGTGCGACCTGTAAAC 240
   |||
Db 61 MetLysGlnLeuGlnProAlaLeuGlnTrpGlyThrGluGluLeuLysSerLeuLysTrp 80
QY 241 ACCGTGGCCACCTGTACTGTCTGTGCAACCGCGCATGTAGATCCGCGACACCAAGAGCC 300
   |||
Db 81 ThrValAlaThrLeuLysCysValIleGlnLysIleGlnValArgAspThrLysGlnAla 100
QY 301 CTGGAACAAGATGAGAGAGAGCAACAAGTCCCGAGAGAAACCCAGAGGCCAGAGAG 360
   |||
Db 101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
QY 361 GCGGACGCGAAGGTGAGCGCAAGACTACCCCATGTGACAGAACTGTGAGGGCCAGATG 420
   |||
Db 121 AlaAspGlyLysValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlnGlnMetVal 140
QY 421 CACCGAGGCTATGACCCCGCGACCCCTGAACGCTGTGGGTGAAGGTATGAGAGAGGCC 480
   |||
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLysAla 160
QY 481 TTCAGCCCGGAGGTATCCCGCATGTTTCAACCGCCCTGAGAGGGGGGCGCAACCCCGAGAG 540
   |||
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsp 180
QY 541 CTGAACACAGATGTGAACAACCGTGGCGCGCACAGCGCGCCATCAGATGCTGAAGAGAC 600
   |||
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlnIleGlnAlaIleMetGlnMetLeuLysAsp 200
QY 601 ACCATCAACAGAGAGCGCGCGAGTGGACCGCGCTGTCAACCCCGTGCAGGCGCGCCGCTG 660
   |||
Db 201 ThrIleAsnGlnGluAlaGlnIleTrpAspArgLeuHisIleProValHisAlaGlyProIle 220
QY 661 GCGCCCGGCGAGATCGCGACCCCGCGGCGAGCATCGCGCGGCGCAACCGAGACCCCTG 720
   |||
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240
QY 721 CAGAGACGATCGCTGTGATGACCAACCGCCCGTGTGGCGCGAGATCTACAG 780
   |||
Db 241 GlnGlnGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleGlyLys 260
QY 781 CGGTGGATCATCTGGGCTTGAACAAGTCTGTGGAGATGACAGCCCGCTGTGACATCTGT 840
   |||
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIleLeu 280
QY 841 GACATCCGCGCAGGGCGCGCAAGAGCGCTTCCGCGCATCTGCGGAGACCGGCTTCTTAAGAC 900
   |||
Db 281 AspIleArgGlnGlnGlyProLysGlnProPheArgAspLysValAspArgPhePheLysThr 300
QY 901 CTGGCGCGCGAGAGCGCACCGACGCTGAAGAACTGGATGACCGAGACCTCTGTGTG 960

```

```

Db 301 LeuArgAlaGlnGlnAlaThrGlnGlnValLysAsnTrpMetThrAspThrLeuVal 320
QY 961 CAGAACGCCAACCCCGCATGTGCAAGACCATCTGTGGCGCTCGGCGCCGCGCGCACCTG 1020
   |||
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340
QY 1021 GAGAGATGATGACCGCTGTGCAAGGCGGTGGAGCGGCGCGCGCGCAACAGCGCGGTGTG 1080
   |||
Db 341 GlnGlnMetMetThrAlaCysGlnGlnValGlyGlyProGlyLysLysValArgValLeu 360
QY 1081 GCGGAGGGGATGAGCGGACCGCAACAGCTGTGAACATCAAGATGCAAGAGACCACTTCAAG 1140
   |||
Db 361 AlaGlnAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 380
QY 1141 GCGCCCGCGCGCAACGTCATGCTTCAACTGTGCGGCAAGAGAGGCGCACATGCCAAGAC 1200
   |||
Db 381 GlyProArgAlaGlyIleValLysCysPheAsnGlyGlyLysGlnGlnIleAlaArgAsn 400
QY 1201 TCCGCGCGCCCGCGCAAGAGGCTGTGAAGTGTGGCGCAAGAGGCGCACAGATGAG 1260
   |||
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlnIleGlnMetLys 420
QY 1261 GACTGACCGGACCGCGCGCAACTTCTGTGGCGCAAGATCTGGCCGCAAGAGGCGCGC 1320
   |||
Db 421 AspCysTrpGluArgGlnAlaAsnPheLeuGlnLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGGCAACTTCTGTGCAACAACCGAGCGAGCGCCCGCGCACCGTGCACCGCGCCCGC 1380
   |||
Db 441 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454
QY 1381 CCGCGCGGAGAGCTTCCGCTTGAAGAGACCAACCCCGCGCGCAAGAGCGCGCAAGAGAC 1440
   |||
Db 455 ProAlaGluSerPheArgPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 474
QY 1441 GCGGAGCCCTTACCGGAGCGCCCTGTGACCGCGCTGTGCGAGCGCTGTGCGAGCGCGCTG 1500
   |||
Db 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490
QY 1501 AGCCAG 1506
   |||
Db 491 SerGln 492

```

RESULT 4  
 US-10-332-413-8  
 ; Sequence 8, Application US/10332413  
 ; Publication No. US2004011660A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnstone, Robert Edward  
 ; APPLICANT: Swanson, Ronald Ivar  
 ; APPLICANT: Morris, Lynn  
 ; APPLICANT: Karim, Salim Abdool  
 ; APPLICANT: Williamson, Carolyn  
 ; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H  
 ; FILE REFERENCE: 4569-281393  
 ; CURRENT APPLICATION NUMBER: US/10/332,413  
 ; PRIOR FILING DATE: 2003-09-12  
 ; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
 ; PRIOR FILING DATE: 2001-07-09  
 ; PRIOR APPLICATION NUMBER: US 60/216,995  
 ; PRIOR FILING DATE: 2000-07-07  
 ; PRIOR APPLICATION NUMBER: ZA 2000/3437  
 ; PRIOR FILING DATE: 2000-07-10  
 ; PRIOR APPLICATION NUMBER: ZA 2000/4924  
 ; PRIOR FILING DATE: 2000-09-15  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 8  
 ; LENGTH: 631  
 ; TYPE: PRT  
 ; ORGANISM: Human immunodeficiency virus type 1  
 US-10-332-413-8



```

DB: 14 Gaps: 2
US-09-475-704a-4 (1-1509) x US-10-339-217-109 (1-492)
QY 1 ATGGGCGCCGCGCGCATCTTGGCGGCGAGAGAGTGGAGAAAGATCCGC 60
Db 1 MetGlyAlaSerIleLeuArgGlyGlyLeuLeuAspTrpGluYsIleArg 20
QY 61 CTGGCGCCGCGCGCGCAAGAACATGCTGAGCACTGGTGTGGCGCGCGAG 120
Db 21 LeuArgProGlyGlyYsYsYsYsYsYsYsYsYsYsYsYsYsYsYsYs 40
QY 121 CTGGAAGGCTTGGCGCTGAAACCCCGCTGAGAGCCGCGAGGCGTGAAGAGATC 180
Db 41 LeuGluArgPheAlaLeuAsnSerGlyLeuLeuGluThrAlaGluGlyCysYsGlnIle 60
QY 181 ATGAACAGCTGAGCGCCGCGCTGAGAGCGCGAGCGAGAGTGGCGAGCTGTACAAC 240
Db 61 IleYsGlnLeuGluThrAlaLeuGlnThrGlyThrGluLeuYsSerLeuYsTrpAsn 80
QY 241 ACCGTGGCCACCTGTACTGCTGTCACGCGCGCATGAGGTCCGCGACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuYsCysValAlaIleGlyIleGluValArgAspThrYsGluAla 100
QY 301 CTGGAACAAGATCGAGAGAGAGCAAGCAAGTCCAGAGAGAGCCAGAGGCGCAAGAG 360
Db 101 LeuAspLysIleGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120
QY 361 GCCGAGCGCAAGGTGAGCGCAAGAACTACCCATGCTGAGAGAGCTGAGAGGCGCATGTG 420
Db 121 AlaAspGlyLeuValSerGlnAsnYsProIleValGlnAsnLeuGluGlnIleMetVal 140
QY 421 CACCGAGCGCATGACCCCGCGACCCCTGAACGCTGGGTGTAAGTATCGAAGAGAGGC 480
Db 141 HisGlnProIleSerProArgThrLeuAsnAlaTrpValIleGluGluYsAla 160
QY 481 TTCAGCCCGAGGTGATCCCATGTTTCACCGCTGAGAGAGAGGCGCGCCAGAGAGC 540
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180
QY 541 CTGAACACGATGTTGAACACCGTGGCGCGCGCACAGCGCGCATGAGATGCTGAAGAGC 600
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlnIleAlaIleMetGlnMetLeuYsAsp 200
QY 601 ACCATCAACGAGAGAGCGCGAGTGGAGCGCGCTGAGACCCGCTGAGAGGCGCGCCCTG 660
Db 201 ThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisProValHisIleGlyProVal 220
QY 661 GCCCGCGCGAGATGCGCGACCCCGCGCGAGCATGCGCGCGCGCGCACAGACCGCTG 720
Db 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrIleu 240
QY 721 CAGGAGCGATGCGCTGATGACAGCAACCCCGCGCTGCGTGGGCGCATCTTAAG 780
Db 241 GlnGluGlnIleThrThrMetThrSerAsnProProValProValGlyAspIleYsTrp 260
QY 781 CGGTGATCATCTGGGCGCTGAACAAGATCGTGGAGAGTACAGCCCGCTGAGATCTGCTG 840
Db 261 ArgTrpIleIleLeuGlyLeuAsnYsIleValArgMetYsSerProValSerIleLeu 280
QY 841 GACATCCGCGAGGCGCCCAAGAGGCTTTCGCGCATAGTGAACCGCTTCTTCAAGAC 900
Db 281 AspIleYsGlnGlyProYsGluProPheArgAspTrpValAspArgPheMetYsVal 300
QY 901 CTGGCGCGCGAGAGCGCAACAGAGAGTGAAGACTGATGACAGCAACCTGTGCTG 960
Db 301 LeuArgAlaGluGlnAlaThrGlnAspValYsAsnTrpMetThrAspThrLeuLeuVal 320
QY 961 CAGAACGCGCAACCCGCGATGCAAGACCATCTGCGCGCTGCGCGCGCGCGCGCACCTG 1020
Db 321 GlnAsnAlaAsnProAspCysYsYsYsYsYsYsYsYsYsYsYsYsYsYsYsYs 340
QY 1021 GAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCGCGCGCGCACAGGCGCGCGTGTG 1080

```

```

Db 341 GlnGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAlaArgValLeu 360
QY 1081 GCCGAGCGATGAGCGCGCAACAGCGTGAACATCATGATGACAGAAAGCAACTTCAAG 1140
Db 361 AlaGluAlaMetSerGlnAlaAsnThrThrAsnIleMetMetGlnYsSerAsnYs 380
QY 1141 GCGCCCGCGCGCAAGTCAAGTCTTCAACTGCGGCAAGAGAGGCGCACATGCCAAGAAC 1200
Db 381 GlyProArgArgThrValYsCysPheAsnYsGlyYsGluGlnYsIleAlaLysAsn 400
QY 1201 TGGCGCGCGCGCGCAAGAGGCGTGGAGAGTGGCGCGCGCGCGCACAGATGAG 1260
Db 401 CysArgAlaProArgLysLysGlyCysTrpYsGlyYsGluGlnYsIleGlnMetYs 420
QY 1261 GACTGACCGAGCGCGCGCAAGCAACTTCTGGGCAAGATCTGAGCGCGCAAGAGGCGCGC 1320
Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlySerIleTrpProSerHisLysGlyArg 440
QY 1321 CCGGCGCACTTCTGCGAGACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1380
Db 441 ProGluAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 454
QY 1381 CCGCGCGAGCTTCCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGCGC 1440
Db 455 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnLysAsp 474
QY 1441 CGCGAGCCCTTACCGGAGCCCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1500
Db 475 -----ArgGluProLeuThrSerLeuYsSerLeuPheGlySerAspProLeu 490
QY 1501 AGCCAG 1506
Db 491 SerGln 492

RESULT 6
US-10-339-217-110
; Sequence 110, Application US/10339217
; Publication No. US2003019648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIOR FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 110
; LENGTH: 492
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-110

Alignment Scores:
Pred. No.: 5,57e-117 Length: 492
Score: 2430.00 Matches: 458
Percent Similarity: 94.82% Conservative: 18
Best Local Similarity: 91.24% Mismatches: 16
Query Match: 85.56% Indels: 10
DB: 14 Gaps: 2

US-09-475-704a-4 (1-1509) x US-10-339-217-110 (1-492)
QY 1 ATGGGCGCCGCGCGCATCTTGGCGGCGAGAGAGTGGAGAAAGATCCGC 60
Db 1 MetGlyAlaSerIleLeuArgGlyGlyLeuLeuAspTrpGluYsIleArg 20

```

```

QY 61 CTGCGCCCGCGGCGCAAGACACTGACTGGAAGACCTGTGTGGCCAGCCGCGG 120
DB 21 LeuAaPProGlyGlyLeuValSerCysTyLeuValHisLeuAlaTrpAlaSerArgGlu 40
QY 121 CTGAGAGGGCTTGGCCCTGAACCCGCGCTGTGAGAGACCGCGAGGGCTGCAAGAGATC 180
DB 41 LeuGluAaGPhSerSerLeuAaPProGlyLeuLeuGluThrSerGluGlyCysGlyGlnIle 60
QY 181 ATGAAGACAGCTGCAAGCCGCGCTGCAAGACCGCGACCGAGAGACTGGCAGCTGTACAC 240
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGlnLeuLysSerLeuValTrpAa 80
QY 241 ACCGTGGCCCACTGTACTGCTGTGCAAGCCGCGCTGTGAGAGTCCGCGACCAAGAGAGCC 300
DB 81 ThrValAlaThrLeuPheCysValHisGluLysIleAlaValArgAaPThrLysGlnAla 100
QY 301 CTGGAACAAGATCGAGAGAGAGAGAGAACAGTCCGAGCAGAGAGAGCCAGAGCCAGAGAG 360
DB 101 LeuAaPProGlyGlyLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
QY 361 GCGGAGCGGAGGTGAGCCAGAACTAACCCCATCTGTGAGAGACCTGAGGCGCGAGATGTG 420
DB 121 AlaAaPProGlyThrValSerGlnAaPProGlyLeuAlaGlnAaLeuGlnGlnGlnMetVal 140
QY 421 CACCGAGGCGCATAGCCCGCGCACTGAAACCGCTGGGTGAGAGTGTGAGAGAGAGAGCC 480
DB 141 HisGlnAlaIleSerProArgThrLeuAaPProGlyValValIleGlnGlnLysAla 160
QY 481 TTCAGCCCGGAGGTGATCCCAATGTTCAAGCCGCTGAGAGAGAGGCGCGCCAGAGAGC 540
DB 161 PheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlnAlaThrProGlnAaP 180
QY 541 CTGAACACAGATGTTGAACAACCGTGGGCGGCGCAACAGCCGCGCATGAGATGTGTAAGAC 600
DB 181 LeuAaPProGlyLeuAaPProGlyLeuAaPProGlyLeuAaPProGlyLeuAaPProGly 200
QY 601 ACCATCAACAGAGAGAGCCCGCGAGTGGAGCCGCTGTGCAACCCGCTGAGGCGCGCGCTG 660
DB 201 ThrIleAaPProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220
QY 661 GCGCCCGGCGAGATGCGGAGCCCGCGGCGAGAGATGCGGCGCGCGCGCGAGAGCCGCTG 720
DB 221 AlaAaPProGlyLeuAaPProGlyLeuAaPProGlyLeuAaPProGlyLeuAaPProGly 240
QY 721 CAGGAGCAGATGCGCTGTGATGAGCAAGCAACCCCGCTGCGGTGGGCGAGATCTTACAG 780
DB 241 GlnGlnGlnGlnIleAlaThrPheThrAaPProGlyLeuAaPProGlyLeuAaPProGly 260
QY 781 CGGTGATCATCTGTGGGCTGTGAACAAGATGCTGGAGATGTACAGCCCGCTGAGATCTG 840
DB 261 ArgTrpIleIleLeuGlnLeuAaPProGlyLeuAaPProGlyLeuAaPProGlyLeuAa 280
QY 841 GACATCCGCGAGAGCCCGCAAGAGCCCTGCGGACTAGTGGAGCCGCTTCTTCAAGAGC 900
DB 281 AaPProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
QY 901 CTGCGCGCGCGAGCAGGCGCAACCGAGAGCTGGAAGACTGTGATGAGCAAGCCGCTGTG 960
DB 301 LeuAaPProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
QY 961 CAGAACGCCCAACCCCGAGCTGCAAGACATCTGTGGGCTGTGCGCCCGCGCGCGAGCCTG 1020
DB 321 GlnAaPProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
QY 1021 GAGGAGATGATGAGCCGCTGCAAGAGCCGCGGAGGCGCGCGCGCGAGAGCCGCGCGCTG 1080
DB 341 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 360
QY 1081 GCGGAGCGCATGAGCCAGGCGCAACCGCTGAACATCATGATGAGCAGAGAGCACTTCAAG 1140
DB 361 AlaGlnAlaMetSerGlnThrAaPProGlyLeuAaPProGlyLeuAaPProGlyLeuAaP 380

```

```

QY 1141 GCGCCCGCGGCGCAAGCTCAAGTCTTCAAGTGGGCAAGAGGCGCATGCGCAAGAAC 1200
DB 361 GlyProAaPProGlyIleLeuValSerCysGlyLysGlnGlyHisLeuAlaTrpAla 400
QY 1201 TGCCGCGCGCCCGCGCAAGAGGCTGTGGAAGTGGCGGAGAGAGCCAGCAGATGAAG 1260
DB 401 CysAaPProGlyLeuValSerCysGlyLysGlnGlyHisGlnMetLys 420
QY 1261 GACTGCAACCGAGCGCGCAAGCACTTCTGTGGCAAGATCTGGCCAGCAGCAGAGCCGCG 1320
DB 421 AaPProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 440
QY 1321 CCGCGCACTTCTGTCAGAACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
DB 441 ProGlnAaPProGlyLeuGlnAaPProGlyLeuAaPProGlyLeuAaPProGlyLeuAa 460
QY 1381 CCGCGCGAGAGCTTCCGCTTCAAGAGAGCAACCCCGCGCGCGCGCGCGCGCGCGCGCG 1440
DB 455 ProAlaGlnSerPheAaPProGlyLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 474
QY 1441 GCGGAGCCCTTACCGGAGCCCGCTGACCGCGCTGTGCGAGCTGTGCGAGCGCGCGCTG 1500
DB 475 -----ArgGlnProLeuThrSerLeuLysSerLeuPheGlySerAaPProLeu 490
QY 1501 AGCGAG 1506
DB 491 SerGln 492

RESULT 7
US-10-339-217-140
; Sequence 140, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechler, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of
; FILE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA
; FILE REFERENCE: 111021.143 (ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140
; LENGTH: 491
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-140

Alignment Scores:
Pred. No.: 1,52e-116 Length: 491
Score: 2421.50 Matches: 461
Percent Similarity: 94.82% Conserved: 15
Best Local Similarity: 91.83% Mismatches: 15
Query Match: 85.26% Indels: 11
DB: 14 Gaps: 3

US-09-475-704A-4 (1-1509) x US-10-339-217-140 (1-491)
QY 1 ATGGGCGCGCGCGCGAGCATCTGTGCGGCGAGAGAGTGAACAAGTGGAGAGATCCGC 60
DB 1 MetGlyAlaArgAlaSerIleLeuAaPProGlyLysLeuAaPProGlyLysLeuAaPProGly 20
QY 61 CTGCGCCCGCGGCGGCAAGAGCACTTCTGTGGAAGCACTGTGTGGCGCGCGCGCGAG 120
DB 21 LeuAaPProGlyGlyLysLeuValSerCysGlyLysGlnGlyHisGlnMetLys 40
QY 121 CTGAGAGGGCTTGGCCCTGAACCCGCGCTGTGAGAGACCGCGAGGGCTGCAAGAGATC 180

```

Db 41 LeuGluArpHeAlaValaenProGlyLeuLeuGluThrAlaGluGlyCysIysGlnIle 60  
QY 181 ATGAAGAGCTGAGCCCGCTGAGACCGGACCGGAGAGCTGGACCTGTACAAC 240  
Db 61 IleIysGlnLeuGlnProAlaLeuGlnThrGlyThrGluIuLeuArpSerLeuYrAn 80  
QY 241 ACCGTGGCCACCTGTACTGGTGACCGCGGATGAGGTCCGGACCAACAGAGGCC 300  
Db 81 ThrValAlaThrLeuYrCysValHisAlaGlnIleGluValArpThrIysGlnAla 100  
QY 301 CTGACACAGATCGAGAGAGAGACAACTATCCACAGACAGACCCAGAGCCAGAG 360  
Db 101 LeuArpArGIIeGluGluGlnGlnAnIysSerGlnGlnIuYrThrGlnGlnAlaenGlu 120  
QY 361 GCGGAGGCAAGGTGAGCCGAACTACCCATGTGTGACAACTGCAAGGCGCATGTGT 420  
Db 121 AlaArpGlyIysValSerGlnAnIyYrProIleValGlnAnIuLeuGlnIyGlnIuVal 140  
QY 421 CACGAGGCTATCAGCCCGGACACCTGTGAAGCGCTGGAGGTGATCGAGAGAGGCC 480  
Db 141 HisGlnAlaIleSerProArGlnThrLeuAnAlaTrpValIysValIleGluGluYsAla 160  
QY 481 TTCAGCCCGGAGGTGATCCCATGTTACCGCCCTGAGGAGGCGCACCCCGCAGAG 540  
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnArp 180  
QY 541 CTGAACACGATGTTGAACACCGTGGGCGGCACAGCCCGCATGTGAGATGCTGAAGAC 600  
Db 181 LeuAnThrMetLeuAnThrValGlyGlnHisGlnAlaMetGlnMetLeuYsArp 200  
QY 601 ACCATCAACGAGAGGCGCGGAGTGGACCGCTGACCGCTGACCGCTGACGCGCGCGG 660  
Db 201 ThrIleAnGluGlnAlaIleGluTrpArpArGlnHisIleProValHisAlaIyProIle 220  
QY 661 GCGCCCGGCAAGTGGCGCAACCCCGCGGACGACATGTCGCGCGCACAGACCTGT 720  
Db 221 AlaProGlyIleMetArGlnIuProArGlySerArpIleAlaGlyThrThrSerSerLeu 240  
QY 721 CAGAGACATGCGCTGATGACAGACACCCCGCTGCGTGGGCGACATCTACAG 780  
Db 241 GlnGluGlnIleAlaIleThrMetThrGlyAnProProValProValGlyArpIleYrYs 260  
QY 781 CGGTGATCATCTGGGCTGACAAAGATCGTGGATGTACAGCCCGCTGAGATCTGT 840  
Db 261 ArgTrpIleIleLeuGlyLeuAnIysIleValArpMetYsSerProValSerIleLeu 280  
QY 841 GACATCCGCGACGGCGCCCAAGAGCCCTTCCGCACTACGTGACCGCTTCTTCAAGC 900  
Db 281 ArpIleIysGlnGlyProIysIuProPheArpArGlyValArpArGlnPhePheYrThr 300  
QY 901 CTGCGCGCGGACAGGCGCACAGGACGTGAAGACTGGATGACCGACCTGTGGTG 960  
Db 301 LeuArGlnIleGluGlnAlaThrGlnArpValIysAnIuTrpMetThrArpThrLeuVal 320  
QY 961 CAGAACGCCAACCCCGATGCAAGACCATCTGCGCTTGGGCGCGCGCGCACCTGT 1020  
Db 321 GlnAnAlaArpProArpCysIysThrIleLeuArGlnIleGluYrProGlyAlaSerLeu 340  
QY 1021 GAGGAGATGATGACCGCTGCGGCGGCGTGGGCGCGCGCGCACAGGCGCGCTGT 1080  
Db 341 GlnIuIuMetThrAlaCysGlnIleGlyAlaGlyIuProSerHisIysValArpValIleu 360  
QY 1081 GCGGAGGAGTGAAGCGGACCAAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140  
Db 361 AlaGlnAlaMetSerGlnThrAnSer---ThrIleLeuMetGlnArpSerHisPheYs 379  
QY 1141 GCGCCCGGCGCAAGTGAAGTGTCTTAATCTGCGCAAGAGGCGCACATGCGCAAGAC 1200  
Db 380 GlyProIysArGlnIleValIysCysArpHeAnIyCysGlyIysGlnIuHisIleAlaIysAn 399  
QY 1201 TGCGCGCGCGCGCAAGAGGCGTGGAGAGCGCGGACAGAGGCGCACCAATGAAG 1260  
Db 400 CysArGlnAlaProArGlyIysGlyCysThrIuYsCysGlyIysGlnIuHisIleMetIys 419

QY 1261 GACTGACCGAGAGCGGCGCAACTTCTGGGCAAGATCTGGGCCAGACGCGGCGCGC 1320  
Db 420 ArpCysThrGlnArGlnAlaAnPheLeuGlyIysIleTrpProSerHisIleYsGlyArG 439  
QY 1321 CCGGCAACTTCTTCAAGAACCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1380  
Db 440 ProGlyAnPheLeuGlnSerArpProGlu-----ProThrAlaPro 453  
QY 1381 CCGGCGGAGGCTTCCGCTGAGAGACACCCCGCGCGCGCGCGCGCGCGCGCGAGC 1440  
Db 454 ProAlaGluSerPheArpPheGluIuThrThrProAlaLeuGlnGlnIyProIysArp 473  
QY 1441 GCGGAGCCCTACCGGAGCCCTGTACCGCCCTGCGACCTGTTGCGAGCGCGCGCTG 1500  
Db 474 -----ArgGluProLeuThrSerLeuArpSerLeuPheGlySerArpProLeu 489  
QY 1501 AGCGAG 1506  
Db 490 SerGln 491  
RESULT 8  
US-10-339-217-143  
; Sequence 143, Application US/10339217  
; Publication No. US20030198648A1  
; GENERAL INFORMATION:  
; APPLICANT: Buechter, Douglas  
; APPLICANT: Hou, Xiaohong  
; APPLICANT: Marlor, Christopher W.  
; APPLICANT: Rice, William G.  
; APPLICANT: Yang, Wengang  
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of  
; FILE REFERENCE: 111021.143 (ACH-US1)  
; CURRENT FILING DATE: 2003-01-09  
; PRIOR APPLICATION NUMBER: US 60/347,369  
; PRIOR FILING DATE: 2002-01-11  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 143  
; LENGTH: 491  
; TYPE: PRN  
; ORGANISM: HIV-1  
US-10-339-217-143  
Alignment Scores:  
Pred. No.: 3,92e-116 Length: 491  
Score: 2413.50 Matches: 456  
Percent Similarity: 95.02% Conservative: 21  
Best Local Similarity: 90.84% Mismatches: 14  
Query Match: 84.98% Indels: 11  
DB: 14 Gaps: 3  
US-09-475-704A-4 (1-1509) \* US-10-339-217-143 (1-491)  
QY 1 ATGGGCGCGCGCGCGCGACACCTTGGCGGCGGAGAACTGGAACAATGGGAGAAATCCGC 60  
Db 1 MetGlyAlaArpAlaSerIleLeuArGlyGlyIysLeuArpIyArpGluYsIleArG 20  
QY 61 CTGCGCGCGCGCGGCAAGACACTACATGCTGAAGCACTGGTGTGGGCGGCGGAG 120  
Db 21 LeuArpProGlyIyGlyIysIyShiYrMetIleYsHisLeuValTrpAlaSerArGlu 40  
QY 121 CTGAGGCGCTTGGCCCTTGAACCCCGGCTGTGAGACCGCGGAGGCTTGAAGCAATG 180  
Db 41 LeuGluArpHeAlaLeuAnProGlyLeuLeuGluThrSerGluIyCysIysGlnIle 60  
QY 181 ATGAAGAGTGAAGCGCGCGCTGCAAGACCGGACCGGAGAGAGTGTGGCACTGTACAAC 240  
Db 61 IleIysGlnLeuGlnProAlaLeuGlnThrGlyThrGluIuLeuArpSerLeuYrAn 80  
QY 241 ACCGTGGCCACCTGTACTGGTGACCGCGGATGAGGTCCGGACCAACAGAGGCC 300

```

Db      |||||||.....:|||||.....
81 ThrValaIaThreutryCyserValaHisaIaGlyIleGluIleArgspHrIySglua 100
Qy      CTGGAACAAGATCGAGAGAGAGCAAGAACATCCACAGAGAACCCAGACAGCCAAAGAG 360
Db      |||||||.....
101 LeuAapLyIleGluIleGluIleGluIleGluIleGluIleGluIleGluIleGluIle 120
Qy      GCGGACGGGAGGTGAGCCAGAACTACCCATCTGTCAGAACCTTGAGAGGCCAGAGGTG 420
Db      |||||||.....
121 AlaAapGlyLyValaSerIleAsnIleProIleValaGlnAsnIleGluIleMetVal 140
Qy      CACAGAGCCATGAGCCCGGACACCTGAGACGCTGGGTGAGGTGATCGAGAGAGAGCC 480
Db      |||||||.....
141 HisGlnIleAsnSerProArgHrIleuAsnIleAlaIleValIleGluIleGluIleVal 160
Qy      TTACAGCCCGGAGGTATCCCATGTTTACCGCTGAGAGGAGGCGCCACCCCGACAGAC 540
Db      |||||||.....
161 PheSerProGluIleIleProMetPheThrIleAsnSerGluIleGluIleAlaThrProGlnAap 180
Qy      CTGAACACAGATGTTGAACAACCTGGGGCGGCGCACAGGCGCCATGCAATGCTGAAGAC 600
Db      |||||||.....
181 LeuAsnThrMetLeuAsnThrValaGlyGlyHisGlnIleAlaMetGlnMetLeuYsAap 200
Qy      ACCATCAACGAGAGAGCGCCGAGTGGAGCCGCTGACACCCCGTGCAGAGCGCGCGCCG 660
Db      |||||||.....
201 ThrIleAsnGluIleAlaIleGluIleAlaIleGluIleAlaIleGluIleAlaIleGluIle 220
Qy      GCGCGCGGCGAGATGGCGGAGCCCGCGGAGGAGATCGCGGCGCGCCACAGACCCCTG 720
Db      |||||||.....
221 AlaProGlyIleGluIleMetArgIleProArgIleYsSerAapIleAlaGlyIleThrIleSerThrIleu 240
Qy      CAGGAGACAGATGCGCTGATGACACAGAACCCCGCTGCGCTGGGCGGACATCTACAG 780
Db      |||||||.....
241 GlnGluIleGlnIleAlaIleThrMetThrGlyAsnProProValaProValaGluIleIleYrLyS 260
Qy      CGGTGATCATCTCTGGGCGCTGAGAACAGATCGTGGGATGATGAACGCGCCGTGAGCATCTG 840
Db      |||||||.....
261 ArgTrpIleIleIleuGluIleuAsnIleValaArgMetYsSerProValaSerIleIleu 280
Qy      GACATCCGCGAGAGCGCCCAAGAGCCCTTCCGAGCTACGTGAGCCGCTTTCTTCAAGAC 900
Db      |||||||.....
281 AapIleYsGlnGlyIleProYsGluIleProPheArgAapTrpValaAapArgPhePheYsThr 300
Qy      CTGCGCGCGGAGAGAGCCCAACAGACCTGAAGAAATGATGATGACCAAGACCTTGGTGG 960
Db      |||||||.....
301 LeuArgGlnIleGluIleAlaThrGlnIleAsnValaIleAsnTrpMetThrAapThrIleuVal 320
Qy      CAGAACGCGCAACCCGAGCTGCAAGACCATCTGCGGCTCTGGCGCCCGGCGGACCCCTG 1020
Db      |||||||.....
321 GlnAsnIleAsnProAspYsIleThrIleIleuArgIleAlaGluIleProGlyIleAsnIleu 340
Qy      GAGGAGATGATGACCGCTGCAAGGCGCTGGGCGGCGCCCGGCGCAAGGCGCGCTGCTG 1080
Db      |||||||.....
341 GlnIleuMetThrAlaCyserGlnIleValaGlyIleProSerHisIleValaIleArgValleu 360
Qy      GCGGAGGCGATGAGCGGCGCAACCGCTGAACATATATATGATGAGAAAGCACTTCAAG 1140
Db      |||||||.....
361 AlaGlnIleAsnSerGlnIleAsnIleAsnIleAsnIleAsnIleAsnIleAsnIleAsn 379
Qy      GCGCGCGCGGCGCAACCGCTGAAGCTTCAACTGCGGCGGAGAGGCGCACTCGGCAAGAC 1200
Db      |||||||.....
380 GlyPheIleYsArgThrValaIleYsCyserPheAsnCyserGlyIleGluIleHisIleAlaArgAsn 399
Qy      TGCCGCGCGCGCGGCAAGAGGCGCTGTCGAAGTGGCGGAGAGGCGGCAAGATGAAG 1260
Db      |||||||.....
400 CyserGlnIleProArgIleYsGlyCyserTrpYsCyserGlyIleGluIleHisIleGlnMetYs 419
Qy      GACTGACCGGAGCGCGCAACTTCTGGGCGCAAGATCTGGCGGCGGCAAGAGGCGCGC 1320
Db      |||||||.....
420 AapCyserThrGlnArgIleAlaAsnIleAsnIleGluIleYsIleTrpProSerHisIleYsGlyArg 439
Qy      CCGGCAACTTCTGAGAGACCGGAGCGCGCGCGCGCGCGCAAGCTGCGGCGCGCGCGC 1380

```

```

Db      440 ProGlyAsnPheIleuGlnAsnArgProGlu-----ProThrAlaPro 453
Qy      1381 CCGCGCGGAGAGCTTCCGCTTTCAGAGAGACACCCCGCGCGCGGAGAGAGAGCCCAAGAGAC 1440
Db      454 ProAlaGlnIleSerPheArgPheGluIleGluIleGluIleGluIleGluIleGluIle 473
Qy      1441 GCGGAGCCCTTACCGGAGCGCGCTGACCGCGCTGGCGGAGCGCTGTCGCGGAGCGCGCTG 1500
Db      474 -----ArgGluIleProIleuThrSerIleuYsSerIleuPheIleYsSerAapProIleu 489
Qy      1501 AGCCAG 1506
Db      490 SerGln 491

RESULT 9
US-10-339-217-105
; Sequence 105, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlot, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of
; FILE REFERENCE: 111021.143 (ACH-US1)
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105
; LENGTH: 508
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-105

Alignment Scores:
Pred. No.: 8,996-115 Length: 508
Score: 2387.00 Matches: 454
Percent Similarity: 92.19% Conservative: 18
Best Local Similarity: 88.67% Mismatches: 26
Query Match: 84.05% Indels: 14
DB: 14 Gaps: 3

US-09-475-704A-4 (1-1509) x US-10-339-217-105 (1-508)
Qy      1 ATGGCGCGCGCGCGGCGGACGATCTGCGCGGCGGAGAACTGGAGAACTGGAGAACTCCGC 60
Db      1 MetGlyAlaArgAlaSerIleIleuArgGlyIleuArgIleuArgIleuArgIleuArgIleuArg 20
Qy      61 CTGGCGCGCGGCGGAGAGAGCACTACATGCTGAAGCACTGTGTGTGGCGGAGCGCGAG 120
Db      21 LeuArgProGlyGlyIleYsIleHisIleYsIleuIleYsIleValaIleTrpIleAsnArgIleu 40
Qy      121 CTGAGAGCGCTTGGCGCTGAACCCCGGCGCTGTCGAGACCGCGGAGCGCGCTGCAAGAGATC 180
Db      41 LeuGlnArgPheAlaIleuAsnProGlyIleuIleGluIleuIleuIleuIleuIleuIleuIleu 60
Qy      181 ATGAAGCAAGCTGCAAGCGCGCGCTGCAAGACCGGCGGAGAGAGAGCTGCGAGCTTCAAC 240
Db      61 MetYsGlnIleuIleProAlaIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleu 80
Qy      241 ACCGAGGCGCAACCTGATCTGCGTGCAGCGCGGCGGATCGAGGTCCGAGACCAAGAGAGCC 300
Db      81 ThrValaIaThreutryCyserValaHisaIleGluIleAsnIleYsValaArgAapHrIySglua 100
Qy      301 CTGGAACAAGATCGAGAGAGAGCAAG-----AAGTCCAGACAGAGAACCCAGAGGCC 354
Db      101 LeuAapLyIleGluIleGluIleGluIleGluIleGluIleGluIleGluIleGluIleGluIle 120

```



QY	355	AAGGAGGCGCGAGCGAGGCGAGGCGAGCACTACCCGATGGTGAGAACTCGAGGCGCGAG	413
Db	121	LySAlAlAlAspGluLulysValSerGlnAsnIyrProIleValGlnAsnLeuGlnGlyGln	140
QY	415	ATGTGTACACAGGCGCATCAGCCCGCGACCCCTGAAGCGCTGGGTGAAGGTGATCGAGAG	474
Db	141	MetValHisGlnAsnIleuSerProAdgThrIleuAsnAlaIrrValLylValIleGluGln	160
QY	475	AAGGCGCTCAGGCGCGAGGTGATCCCGAGTTTCAACCGGCTTGAGCGAGGCGCGACCCCG	534
Db	161	LySAlAlPheSerProGluValIleProMetPheThrAlaIleuSerGluLulAlaIrrPro	180
QY	535	CAGGACTTGAACAGCATGTGTGAACAACCTGGGGGGCGCACAGGCGCGCATGAGATGTGTG	594
Db	181	GlnAspLeuSerThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeu	200
QY	595	AAGGACACCATCAAGAGAGGCGCGCGATGGAGACCGCTTGACCCCGCTGACAGCGCGCG	654
Db	201	LySAspThrIrrIleAsnGlnLulAlaIleGluIrrPAspArgLeuIrrProValHisIleGly	220
QY	655	CCCGTGGCGCGCGCGACATGGCGCGACCCCGCGGAGAGACATGGCGCGCGCGACCGAG	714
Db	221	ProMetAlaProGlyGlnIleuArgLulProArgGlySerAspIleAlaGlyThrThrSer	240
QY	715	ACCTGTGACGAGACAGATGCGCTGTGATGACACAGCAACCCCGCGTGCCTGGGCGCAATC	774
Db	241	ThrIleuArgGlnGlnIleAlaIrrPheMetThrSerAsnProIrrProValGlyAspIrrle	260
QY	775	TACAGCGGTGGATCATCTGGGCGCTGGAACAAGATCGTGCGGATGTACAGCCCGCTGAGC	834
Db	261	TyrIrrAspGlyPrrIleIleLeuGlyIleuAsnLylIleValArgMetLylSerIrrValSer	280
QY	835	ATCTGTGACATCCGCGCAGGCGCGCCCAAGAGCCCTTCGCGACTAGCGACCGCTTCTTC	894
Db	281	IleIleuAspIrrIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPhePhe	300
QY	895	AAGACCCCTGCGCGCGCGAGAGCGCACCCAGAGCGGAGAACTGTGATGACGAGACCCCTG	954
Db	301	LySAlAlAspAlaArgAlaGlnGlnIrrAlaThrGlnAspValLylAsnIrrPheMetThrGluThrIleu	320
QY	955	CTGTGTGCAAGAACGCCAACCCCGCATGTGCAAGAACCATCTCGCGCGCTCGCGCGCGCGCC	1014
Db	321	LeuValGlnAsnAlaAsnProAspPrrLylThrIrrIleuLylAlaIleuGlyIrrIleGlyAla	340
QY	1015	ACCGTGGAGGAGATGATGACCGGCTTGGCCAGGCGCGTGGGCGCGCGCGCGCACAGGCGCGC	1074
Db	341	ThrIleuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLylAspAlaArg	360
QY	1075	GTGCTGTGCGGAGCGCATGAGCGCGACGCGCAACCGTGAACATCATATGATCAAGAGAGCAAC	1134
Db	361	ValIleuAlaGluLulAlaMetSerGlnAlaAsnIrrAsnThrAsnIrrIleMetMetGlnIrrSerAsn	380
QY	1135	TTTCAAGGCGCGCGCGCGCAGTCAAGTCAAGTCTTCAACTGCGCGCAAGAGAGGCGCACTCGCC	1194
Db	381	PheLylSerSerLylLylArgIrrIleValLylLylCysSerAsnLylGlyGlnGlyHisIrrIleAla	400
QY	1195	AAGAACTGCGCGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGCAGGAAGAGAGCGCACAG	1254
Db	401	ArgAsnLylAspAlaProArgLulLylSerLylCysIrrLylLylSerGlyLylSerGlnGlyHisGln	420
QY	1255	ATGAAGACCTGCACCGAGCGCGCAGCGCAACTTCTGGGCGAGATCTGGCGCGCACAGACAG	1314
Db	421	MetLylAspPrrCysIrrGlnArgGlnAlaAsnIrrPheLylGlyLylIrrProSerHisLylLyl	440
QY	1315	GGCGCGCGCGCGCACTTCTGTGCAACAACCGACGAGACCGCGCGCGCGCGCACCGCTG	1368
Db	441	GlyIrrArgProGlyAsnIrrPheLylGlnAsnArgProGluIrrProThrAlaProIrrAlaGluSer	460
QY	1369	-----CCCAACCGCGCGCGCGCGCGCGCGAGAGCTTCGCTTGAAGAGAC	1410
Db	461	PheIrrArgAsnArgProGluIrrProThrAlaIrrProIrrAlaGluSerPheIrrArgPheGlnGluThr	480
QY	1411	ACCCCGCGCGCGCGCAAGCAGAGCCCAAGAGCCGCGAGCCCTTACCGCGAGCCCGCTTACCGCC	1470

DB	Accession	Protein Name	Length	Score	Alignment	Score	Length	Matches	Conservative	Mismatch	Indel	Gaps
Db	481	ThirProthPrblysgInuProlysasp	-----	ArgaspProleuthrSer	496							
Qy	1471	CTGGCGAGCCTGTTCGGAGCGGCCCTGAGCGAG	1506									
Db	497	LeuylSerLeuPheglySerAspProSerSerGln	508									
RESULT 10												
US-10-339-217-104												
/ Sequence 104, Application US/10339217												
/ Publication No. US20030198648A1												
/ GENERAL INFORMATION:												
/ APPLICANT: Buechter, Douglas												
/ APPLICANT: Matlor, Christopher W.												
/ APPLICANT: Rice, William G.												
/ APPLICANT: Yang, Wengang												
/ TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of												
/ TITLE OF INVENTION: Nucleoside 7 Protein to HIV-1 RNA												
/ FILE REFERENCE: 111021.143(ACH-US1)												
/ CURRENT APPLICATION NUMBER: US/10/339,217												
/ PRIOR FILING DATE: 2003-01-09												
/ PRIOR APPLICATION NUMBER: US 60/347,369												
/ NUMBER OF SEQ ID NOS: 144												
/ SOFTWARE: FastSeq for Windows Version 4.0												
/ SEQ ID NO: 104												
/ LENGTH: 491												
/ TYPE: PRT												
/ ORGANISM: HIV-1												
US-10-339-217-104												
Alignment Scores:												
Pred. No.:	1,07e-114	Length:	491									
Score:	2385.50	Matches:	453									
Percent Similarity:	93.82%	Conservative:	18									
Best Local Similarity:	90.24%	Mismatch:	20									
Query Match:	84.00%	Indel:	11									
DB:	14	Gaps:	3									
US-09-475-704A-4 (1-1509) x US-10-339-217-104 (1-491)												
Qy	1	ATGGGCGCGCGCGCCAGATCTCTGGCGCGCGCGAAGCTGAGCAAGTGGAGAAAGATCCGC	60									
Db	1	MetGlyAlaArgAlaSerIleLeuArgIleGlyLeuAspIleTyrGlnIleSer	20									
Qy	61	CTGGCGCGCGCGCGCGAAGCACTACATGTCGGAAGCACTGGTGTGGGCGCGCGAG	120									
Db	21	LeuArgProGlyGlyArgIleHisTyrMetLeuIleValIlePheSerArgIle	40									
Qy	121	CTGAGAGGCTTCGCGCTTGAACCCCGCGCTGTGTGAGACCGCGCGAGGCTGCAAGCATC	180									
Db	41	LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnIleCysIleGlnIle	60									
Qy	181	ATGAAGCAGCTGAGCGCGCGCTGCAAGCGGCAACCGAGAGGCTGGCAGCTGTACAC	240									
Db	61	IleuysGlnLeuHisProAlaLeuIleThrIleThrGlnIleValGlySerLeuIleThrAsn	80									
Qy	241	ACCGTGGCCACTGTACTGTGCGTCAAGCGCGCGCTGAGGTCCGCGACACCAAGAGGCC	300									
Db	81	ThrValAlaThrLeuIleCysValIleGlnLeuIleGlnValAlaArgPheThrIleGlnIle	100									
Qy	301	CTGGAACAAGATCGAGAGGAGGAGCAACAACGATCCCGAGCAAGAACCCAGAGGCCAAGAG	360									
Db	101	LeuAspIleIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle	120									
Qy	361	GCCGACGCGCAAGGTGAGCCAGAACTACCCCTCTGAGCAACTGACAGGCGCAGATGGTG	420									
Db	121	AlaAspIleGlnIleValSerGlnIleThrIleValIleGlnIleGlnIleGlnIleGlnIle	140									
Qy	421	CACCAAGCCATCAGCCCCCGCACCTTGAAGCGCTGGGTGAAGGTGATTCAGAGAGAGGCC	480									
Db	141	HisGlnIleIleSerProArgThrLeuAsnAlaIleTyrValIleG										



481 TTCAGCCCGAGGTGATCCCATGTTCAACCCGCTTGAAGGAGGCGCCACCCCGCAGAC 540  
161 PheSerProGluValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsp 180  
541 CTGAACACGATGTTGAACAACCGTGGGCGGCGCACAGGCGCCATGAGATGCTGAAGAC 600  
181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
601 ACCATCAACGAGAGGCGCCCGAGTGGAGCCGCTTGACCCCGCTGACGCGCGCGCGTGG 660  
201 ThrIleAsnGlnGluAlaIleGluTrpAspArgLeuHisProValHisIleGlyProAla 220  
661 GCGCCCGGCGAGATGGCGCAACCCCGCGCGAGAGATGGCGCGCGCCACGACACCTGG 720  
221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
721 CAGAGACGATGCGCTGATGACACAGCAACCCCGCTGCGCGCGCGAGCATCTTCAAG 780  
241 GlnGlnGlnIleAlaIleTrpMetThrGlyAsnProProValProValGlyAspIleTrpLys 260  
781 CGGTGATCATCTGGGCTTGAACAAGATCTGCGGATGATGACCGCTTCTTCAAGACC 840  
261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIleLeu 280  
841 GACATCCGCGAGGCGCCCAAGAGCCCTTCCCGATAGTGAACCGCTTCTTCAAGACC 900  
281 AspIleLysGlnGlyProLysGluProPheArgAspTrpValAspArgPheProLysVal 300  
901 CTGCGCGCGAGACAGGCGCACACGAGACGTGAAGATGATGACACGAGACCTGCTGCTG 960  
301 LeuArgIleGlnGlnIleAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 320  
961 CAGAACGCCAACCCCGATGCAAGACCATCTGCGGCTTCCGCGCGCGCGCGCACCTGG 1020  
321 GlnAsnAlaAsnProAspCysIleThrIleLeuLysAlaLeuGlyProAlaIleSerLeu 340  
1021 GAGGAGATGATACCGCTTCCGAGGCGCTGGCGCGCGCGCGCACAGGCGCGCTGCTG 1080  
341 GlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 360  
1081 GCGGAGCGATGAGCGGCGCAACGCTGAACATCATGATGACAGAGAGCAACTTCAG 1140  
361 AlaGlnAlaMetSerGlnAlaAsnSer---AsnIleMetMetGlnArgSerAsnPheLys 379  
1141 GCGCCCGCGGCAACGTCAGTCTTCAACTGCGGCGAGAGGCGCACATCGCCCAAGAC 1200  
380 GlySerLysArgIleValLysCysPheAsnGlyGlyGlnGlyHisIleAlaArgAsn 399  
1201 TCCCGCGCGCGCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCACAGATGAAG 1260  
400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGlnGlyHisGlnMetLys 419  
1261 GACTCAGCGAGCGCGCAACTTCTGAGCAAGATCTGGCGCGAGCGCGCACAGAGGCGCG 1320  
420 AspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439  
1321 CCGGCAACTTCTGAGAACCGCGAGGAGCGCGCGCGCGCACCGTGGCGCACCGCGCG 1380  
440 ProGlyAsnPheLeuGlnSerArgProGlu-----ProThrAlaPro 453  
1381 CCGCGCGAGAGTTCGCTTCCGAGAGACACACCCCGCGCGCGCACAGAGAGCGCGAGAC 1440  
454 ProAlaGlnLysPheArgPheGlnGlnGlnTrpProAlaProLysGlnLysLysAsp 473  
1441 CGCGAGCCCTTACCGAGCGCGCTGACCGCGCTGCGAGCGCTTTCGCGAGCGCGCGCTG 1500  
474 -----ArgGluProLeuLysSerLeuPheGlySerAspProSer 489  
1501 AGCCAG 1506  
490 SerGln 491

RESULT 11  
US-10-339-217-107  
; Sequence 107, Application US/10339217  
; Publication No. US20030198648A1  
; GENERAL INFORMATION:  
; APPLICANT: Buechter, Douglas  
; APPLICANT: Hou, Xiaohong  
; APPLICANT: Marlor, Christopher W.  
; APPLICANT: Rice, William G.  
; APPLICANT: Yang, Wengang  
; TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of  
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA  
; FILE REFERENCE: 111021.143(ACH-US1)  
; CURRENT APPLICATION NUMBER: US/10/339,217  
; PRIORITY FILING DATE: 2003-01-09  
; PRIOR FILING DATE: 2002-01-11  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 107  
; LENGTH: 497  
; TYPE: PRT  
; ORGANISM: HIV-1  
US-10-339-217-107

Alignment Scores:  
Pred. No.: 3,5e-114 Length: 497  
Score: 2375.50 Matches: 451  
Percent Similarity: 93.04% Conservative: 17  
Best Local Similarity: 89.66% Mismatches: 28  
Query Match: 83.64% Indels: 7  
DB: Gaps: 3

US-09-475-704A-4 (1-1509) x US-10-339-217-107 (1-497)

1 ATGGGCGCGCGCGCGCGCATCTGCGCGCGAGAGCTGGAACAAGTGGAGAGATCCGC 60  
1 MetGlyAlaArgAlaSerIleLeuSerGlyGlyLysAspLysTrpGlnArgIleArg 20  
61 CTGCGCGCGCGCGCGCAAGAGCATACATGCTGAAGACCTGTGTGGCGACCGCGGAG 120  
21 LeuArgProGlyGlyLysLysIleTrpMetLeuLysIleValIleTrpAlaSerArgGlu 40  
121 CTGAGAGGCTTGGCGCTTGAACCCCGCTGCTGAGAGACCGCGAGGCTGCAAGCATC 180  
41 LeuGlnArgPheAlaLeuAsnProGlyLeuGlnGlnTrpAlaGlnGlyCysLysGlnIle 60  
181 ATGAAGACGCTGACCGCGCGCTGCGAGACCGGACCGGAGAGCTGCGAGCGCTGTACAC 240  
61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnLeuArgSerLeuPheAsn 80  
241 ACCGTGCGCACCTGTACTGCTGTCAGCGCGCGCATCGAGTCCGCGACACCAAGAGGCC 300  
81 ThrValAlaThrLeuTrpCysValHisLysGlyIleGluValArgAspThrLysGlnAla 100  
301 CTGGAACAAGTGAAGAGAGAGCAACAAGTCCCAAGCAAGAACCCAGAGCGCGCAAGAG 360  
101 LeuAspLysIleGlnGlnGlnGlnAsnLysCysGlnGlnLysAlaGlnGlnAlaLysAla 120  
361 GCGGAGCGAGAGTGAAGCAGAACTAACCCCATCTGCGAGAACTGCGAGGCGCAGATGGTG 420  
121 AlaAspGluLysValSerGlnAsnTrpProIleValGlnAsnAlaGlnIleMetVal 140  
421 CACGAGCCATCAGCGCGCGCACCTGGAACGCTGGTGAAGTATGAGAGAGAGCGC 480  
141 HisGlnIleIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLysAla 160  
481 TTCAAGCCCGAGAGTATCCCATGTTCAACCGCGCTGACGAGAGGCGCGCACCCCGCAGAC 540  
161 PheAsnProGluValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsp 180  
541 CTGAACACGATGTTGAACAACCGTGGCGGCGCACGAGCGCGCATGAGTGTGAAGAGAC 600

```

Db      181 LeuAanthrMetLeuAanthrValGlyGlyHisGlnAlaAlaMetGlnMetLeuValAsp 200
QY      601 ACCATCAAGAGAGAGCCCGGAGTGGAGCCGCTGACCCCGCTGACAGCGCGCCGCTG 660
Db      201 ThrIleAsnGlnGlnIleAlaGlnTrpAspArgThrHisProValHisAlaGlyProVal 220
QY      661 GCGCCCGGCGAGATGGCGAGCCCGCGGCGAGAGATGCGCGCGCGCCCGCCACGACCCG 720
Db      221 AlaProGlnIleMetArgGlnProAlaGlnSerAspIleAlaGlnThrThrSerThrLeu 240
QY      721 CAGAGCAGATGCGCTGATGACCAAGACCCCGCGTGGCGCGCTGGCGCAGATCTTCAAG 780
Db      241 GlnGlnGlnIleAlaTrpMetThrSerAsnProIleProValGlnAspIleTyrLys 260
QY      781 CGGTGATCATCTGGGCTTGAACAAGATGTCGGATGTACAGCCCGCTGACATCTTG 840
Db      261 ArgTrpIleIleLeuIleLeuIleValLeuValArgMetLysSerProValSerIleLeu 280
QY      841 GACATCGCGCAGGCGCCCAAGAGCCCTTCCGCGCATACGTGGACCGCTTTCAGAGCC 900
Db      281 AspIleLysGlnIleProLysGlnProIleArgAspTyrValAspArgPhePheLysTrp 300
QY      901 CTGCGCGCGCAGAGCAGCCCAAGAGCTGAAGACTGTGACCGAGACCTGCTGTG 960
Db      301 LeuArgAlaGlnGlnIleAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320
QY      961 CAGAACGCCAACCCCGACTGCAAGACATCTCGCGCTGCGGCGCCGCGCCACCTG 1020
Db      321 GlnAsnAlaAsnProAspLysTrpIleLeuAlaGlnLeuGlyProIleValSerLeu 340
QY      1021 GAGGAGATGATGACCGCTGCGAGGCGTGGCGCGCGCCGCGCAGAGCCGCGTGTG 1080
Db      341 GlnGlnMetMetThrAlaCysGlnGlnValGlyGlyProSerHisLysAlaArgValLeu 360
QY      1081 GCCAGAGCGATAGCCAGCCCAAGCGTGAACATCATGATGATGACAGAGCAACTTCAG 1140
Db      361 AlaGlnIleMetSerGlnThrAsnSer--AsnIleLeuValGlnArgSerAsnPheLys 379
QY      1141 GCGCCCGCGCGCAAGTCAAGTTCATCTGCGCGCAGAGGCGCAGATCCCGCAAGAC 1200
Db      380 GlySerAsnArgIleValLysCysPheAsnCysGlyLysValGlyHisIleValArgAsn 399
QY      1201 TGC CGCGCGCGCGCAGAGAGGCTGTGGAAGTGGCGCAGAGAGGCGCAGATGAG 1260
Db      400 CysAlaGlnProArgLysGlyCysTrpLysCysGlyGlnGlnGlyHisGlnMetLys 419
QY      1261 GACTGACCGAGCGCGCAACTTCTGGCGCAAGATTTGGCCCAAGCCCAAGAGGCGCGC 1320
Db      420 AspCysTrpGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439
QY      1321 CCGCGCAACTTCTCTCAGAACCGCAGCGCGCGCGCGCGCCGCGCGT--CGCAGCGCGC 1377
Db      440 ProGlnAsnPheLeuGlnAsnArgProGlnProThrAlaProProAlaGlnProThrAla 459
QY      1378 CCGCGCGCGCAGAGCTTCCGCTTCAAGAGACACCCCGCGCCCGCAGAGAGCGCCAG 1437
Db      460 ProProAlaGlnSerPheArgPheGlnGlnThrThrProValProLys----- 475
QY      1438 GACCGCGAGCCCTTACCGGAGAGCCCTGACCGCGCTGCGCAGGCTTGTGGCAGCGCGCC 1497
Db      476 ---ArgGlnLysGlnArgGlnProLeuThrSerLeuLysSerLeuPheGlyAsnAspPro 494
QY      1498 CTGAGCCAG 1506
Db      495 SerSerGln 497

RESULT 12
US-10-339-217-106
; Sequence 106, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong

```

---

```

; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; TITLE OF INVENTION: Nucleosopside 7 Protein to HIV-1 RNA
; FILE REFERENCE: 111021.143(ACh-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; PRIORITY FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106
; LENGTH: 487
; TYPE: PRP
; ORGANISM: HIV-1
US-10-339-217-106

Alignment Scores:
Pred. No.: 4.72e-113 Length: 487
Score: 2353.50 Matches: 448
Percent Similarity: 94.04% Conservative: 25
Best Local Similarity: 89.07% Mismatches: 13
Query Match: 82.87% Indels: 17
Db: 14 Gaps: 5

US-09-475-704a-4 (1-1509) x US-10-339-217-106 (1-487)
QY      1 ATGGCGCGCGCGCGCAGATCTTGGCGCGCAGAGAGCTGCAAGTGGAGAGATCCGC 60
Db      1 MetGlnAlaAlaGlnAlaSerValLeuLysGlnLysLysLeuAspTrpGlnArgIleArg 20
QY      61 CTGCGCGCGCGCGCAGAGACCACTACATGCTGAAGCACTGTGTGTGGCGCGCGCGAG 120
Db      21 LeuArgProGlnGlyLysLysHisIleTyrMetLeuLysHisLeuValTrpAlaSerArgGln 40
QY      121 CTGAGGCGCTTGGCGCTGGAACCCCGCGCTGTGAGACCGCGCGAGGCTGCAAGCATC 180
Db      41 LeuArgPheAlaLeuAsnProGlnLysLeuLeuGlnTrpAlaGlnGlyCysLysGlnIle 60
QY      181 ATGAGACAGCTGCAAGCCCGCTGCAAGCCGCGCAGAGAGTGTGCGAGCTGTACAAAC 240
Db      61 MetGlnGlnLeuGlnSerAlaLeuGlnThrGlyThrGlnGlnLeuArgSerLeuTyrAsn 80
QY      241 ACCGTGCGCACCTCTGTACTGTGTGTCAAGCGCGCATCGAGTCCGCGACCAAGAGGCC 300
Db      81 ThrValAlaThrLeuTyrCysValHisLysGlnIleAspValArgAspTrpLysGlnAla 100
QY      301 CTGACAAAGATGAGAGAGCAGAACAGATCCCGCAGAGAACCTGCAAGGCGCGAGATG 360
Db      101 LeuAspLysIleGlnGlnGlnGlnAsnLysSerGlnLysTrpGlnGlnAlaGlnAla 120
QY      361 GCGGAC--GCGAAGTGAAGCAGAACTACCCCATCTGTGCAAGCTGCAAGGCGCGAGATG 417
Db      121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyMet 140
QY      418 GTGACAGAGCATCACCGCCCGCGACCCCTGAAGCGCTGTGAGAGTGTGAGAGAGAG 477
Db      141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLys 160
QY      478 GCGTTGAGCCCGAGGTATCCCATGTTCACCGCTGTGAGAGAGGCGCGACCCCGCAG 537
Db      161 AlaPheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlyAlaTrpProGln 180
QY      538 GACTTGAACAGATGTTGAACACCGTGGCGCGCGCAGCGCCGCGCATGACATGCTGAAG 597
Db      181 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLys 200
QY      598 GACACCATCAAGAGAGAGCGCGCGAGTGGAGACCGCTGCAAGCCCGCTGACAGCGCGGCC 657
Db      201 AspThrIleAsnGlnGlnAlaAlaGlnTrpAspArgLeuHisProValHisAlaGlyPro 220
QY      658 GTGGCGCGCGCGCAGATCGCGCAGACCCCGCGGAGCGACATCGCGCGCGCACCGAGCAC 717

```

Db 221 ILEAIPROGlyGIMetArGluPProArGlySerAspIleAlaGlyThrThrSerThr 240  
QY 718 CTGCAGAGCAGATCGCTTGATGATGACCAAGCAACCCCCGTCGCCGCTGGGCGCAATCTAC 777  
Db 241 LeuGInGluGlnIleAlaTrpMetThrSerAsnProProlIleProValGlyAspIleTyr 260  
QY 778 AAGCGGTGATATCATCTGGGGCCCGGAACAAGATCGTCCGATATGACAGCCCCCGGAGAGATC 837  
Db 261 LySAATGTPPLeIleLeuGlyLeuAanLySIIeValArGmetTyrSerProValSerIle 280  
QY 838 CTGCAGATCCGCGAGGGCCCCAAGAGAGCCCTTCGCGACTACGTGAGCGGCTTCTCAAG 897  
Db 281 LeuAspIleLeuGlnGlyProlySGluPProPheArAspTyrValAspAlaGpPheLeuS 300  
QY 898 ACCCTGCGCGCCGAGAGGCCAACCCAGAGCGTGAAGAACTGATGACCGAGACCTTCGCTG 957  
Db 301 ThrLeuArSAlaGluGlnSerSerGInGluValIyAsnTrpMetThrAspThrLeuLeu 320  
QY 958 GTGCAGAAACGCCAACCCCGCATCTCAAGACCATCTTCGCGCTTCGCGCCCCGCGCAC 1017  
Db 321 ValGInAsnAlaAsnProAspCysIyTrIleLeuArGAlaLeuGlyProGlyAlaThr 340  
QY 1018 CTGAGAGAGATGATGACCGCTGCGCGAGGGGCGTGGGCGGCCCGGCGCAAGAGCGCGGTG 1077  
Db 341 LeuGInGluMetMetThrIlaCysGInGlyValGlyIProGlyIlySlyValAlaArgVal 360  
QY 1078 CTGCGCGAGGCGCATGAGCGAGGCCAACAGCGTGAACATCATGTATGAGAGAAGCAATTTC 1137  
Db 361 LeuAlaGluAlaMetSerGInIlaAanThr---AsnIleMetMetGInlySsSerAsnPhe 379  
QY 1138 AAGGCGCCCCCGCGCGCAACGTCAATGCTCTCACTCGCGGCAAGAGGGCCACATCGCCAG 1197  
Db 380 LysGlyProLyArGThrValIySsCysPheAsnCysGlyLySGluGlyAsIIeAlaArg 399  
QY 1198 AACTGCGCGGCCCGCCGCAAGAGGGCTCTGTGAAGTGGGGCAAGAGGGGCCACCGATG 1257  
Db 400 AsnCysArGAlaProArGlySlySGlyCysTrpLyCysGlyLySGluGlyAsIImet 419  
QY 1258 AAGACTGACCGAGCGCGAGGCCCACTTCCTGGGCAAGATCTGGCCCGCAAGAGGCG 1317  
Db 420 LysAspCysThrGluArGInIlaAanPheLeuGlyLyAlaIleTrpProSerTyrIyLySGly 439  
QY 1318 CGCCCCGGCACTTCTCTGCAAGACCGCAGCGAGAGCCCGCGCCCAACCGTCCACCGCC 1377  
Db 440 ArgSerGlyAsnPheLeuGInSerArGProGlu-----ProSerAla 453  
QY 1378 CCCCCCGCGAAGCTTCGCTTCGAGAGAGACCAACCCCGCCCGCCCAAGAGGCCCAAG 1437  
Db 454 ProProlaGlnSerPheArGpRegGluGlu-----ArgGluProLyS 467  
QY 1438 GACCCGAGAGCCCTTACCGCGAGCCCGCTGACCGCCCTGCGAGCGCTGTTGGCAGCGGCC 1497  
Db 468 AspLySGluPro-----ProLeuThrSerLeuLySerIleuPheGlySerAspPro 484  
QY 1498 CTGAGCCAG 1506  
Db 485 SerSerGln 487

RESULT 13  
US-10-339-217-108  
; Sequence 108, Application US/10339217  
; Publication No. US20030198648A1  
; GENERAL INFORMATION:  
; APPLICANT: Buechter, Douglas  
; APPLICANT: Hou, Xiaohong  
; APPLICANT: Marlor, Christopher W.  
; APPLICANT: Rice, William G.  
; APPLICANT: Yang, Wengang  
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of  
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA  
; FILE REFERENCE: 111021.143(ACH-US1)  
; CURRENT APPLICATION NUMBER: US/10/339, 217

```

CURRENT FILING DATE: 2003-01-09
PRIORITY FILING DATE: 2002-01-11
PRIORITY FILING DATE: 2002-01-11
NUMBER OF SEQ ID NOS: 144
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 108
LENGTH: 496
TYPE: PRT
ORGANISM: HIV-1
US-10-339-217-108

Alignment Scores:
Pred. No.: 9,056-113 Length: 496
Score: 2348.00 Matches: 447
Percent Similarity: 92.11% Conservative: 20
Best Local Similarity: 88.17% Mismatches: 24
Query Match: 82.68% Indels: 16
DB: 14 Gaps: 4

US-09-475-704A-4 (1-1509) x US-10-339-217-108 (1-496)

QY 1 ATGGGCGCCCGGCGGCGAGCATCTCTGGCGGCGGAGAAAGCTGGACAATGGGAGAAAGATCCGC 60
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyIysLeuAspLysTrpGlnYsIleArg 20

QY 61 CTGGCGCCCGGCGGCGGAGAACATCAATGCTGAAGACACTGGTGTGGCGCGCGAG 120
DB 21 LeuArgProGlyGlyIysIysIysTrpMetLeuYsHisLeuValTrpAlaSerArgGlu 40

QY 121 CTGGAGGCGCTTCGCGCTGAACCGCGCGCTGTGAGACCGCGGAGGCTGCAAGCAGATC 180
DB 41 LeuGluArgPheAlaLeuAsnSerGlyLeuLeuGluThrSerAspLysIysGlnIle 60

QY 181 ATGAAGCAGCTGCGACCGCGCTCTGACAGACCGGACACCGAGAGACTGGCGAGCTGTACAC 240
DB 61 ILeuGlnIleuGlnProAlaLeuIysThrGlyThrGlnIleuArgSerLeuTrpAsn 80

QY 241 ACCGTGGCACCCCTGTACTCGGTGACGCGCGGATCGAGTCCGCGGACCAAGAGAGGCC 300
DB 81 ThrValAlaThrLeuTrpCysValHisAsnAsnIleGluIleArgAspThrIysGlnAla 100

QY 301 CTGGAAGATCGAGAGAGAGAGAGAAACAATGCCACAGAGAACCCAGAGGCCAAG-- 357
DB 101 LeuAspArgIleGlnIleGlnIleGlnIleIysIysCysGlnIleGlnIleThrGlnIleGlnIle 120

QY 358 -----GAGCGCGAGCGGCAAGGTGAGACCGAATCACTCCCATCTGTGCAAGACCTG 405
DB 121 GlnGlnValGlnAlaAlaAspGlyIysValSerGlnAsnTrpProIleValGlnAsnLeu 140

QY 406 CAGGGCGCATGTGATGATCAACAGGCGCATGACCGCGCGACCTGTAAAGCTGGGAGAGTG 465
DB 141 GlnGlyGlnMetValHisGlnSerLeuSerProAspGlnIleuAsnHisAspValIysVal 160

QY 466 ATCGAGAGAAAGCGCTTCAGCGCGCGAGGTATCCCATGTTTCAACGCGCTGTGACGAGGC 525
DB 161 ILeuGlnIleuYsAlaPheSerProGluIleIleProMetPheThrAlaLeuSerGlnGly 180

QY 526 GCCACCGCGCGGAGCTGTAAACAATGTTTAAACACCTGTGGGCGGCGCAACAGGCGCGCATG 585
DB 181 AlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyIysHisGlnAlaAlaMet 200

QY 586 CAGATGCTGAAGAGACACCATCAACAAGAGAGCGCGCGAGTGGGAACGCTGTGACCGCGTG 645
DB 201 GlnMetLeuYsAspThrIleAsnGlnGlnAlaAlaGlnTrpAspArgLeuHisProVal 220

QY 646 CAGGCGGCGCGCGTGGCGCGCGCGCGAGATGCGGACCGCGCGGAGAGACATCGCGGC 705
DB 221 HisAlaGlyProValAlaProGlyGlnMetArgGlnProArgGlySerAspIleAlaGly 240

QY 706 GCCACCGAGACCTCTGAGAGAGAGAGATGCCCTGAGAGACAGAGAACCGCGCGCGCGG 765
DB 241 ThrThrSerAsnLeuGlnGlnIleGlnIleAsnTrpMetThrAlaAsnProIleProVal 260

```





```
QY 1009 GGGGCCA CCTGGAGGAGATGATACCGCTGCGCCAGGGCGTGGGGCCCCGGCCACAG 1068
    |||||
Db 340 GtYAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerH1SLys 359
    |||||
QY 1069 GCCGCGCTGCTGGCGGAGCGGATAGCGCAGCCACAGCGTGAAC-----ATCATGATG 1122
    |||||
Db 360 AlArGValLeuAlaGluAlaMetSerGlnAlaThrAsnLeuAsnGlyAlaAlaMetMet 379
    |||||
QY 1123 CAGAAAGCAACTTCAAGGGCCCCCGGCGCAACGTCAAGTCTTCAACTGGGCAAGAG 1182
    |||||
Db 380 GlArGSerAsnPhelYsglyProArGlyS11eVallyeCysPhaenCysGlyLysGlu 399
    |||||
QY 1183 GGCACATGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGGCTGCTGAAAGTGGCGCAG 1242
    |||||
Db 400 GlyH1s11eAlaArgAsnCysLysAlaProArGlyLysGlyCysTrpLysCysGlySer 419
    |||||
QY 1243 GAGGGCCACCAGATGAAGACTGCAACCGAGCGCCAGGCGCAACTTCTGGGCAAGATCTGG 1302
    |||||
Db 420 GluGlyH1sGlnMetLysAspCysThrGluArGlnAlaAsnPhelLeuGlyLys11eTrp 439
    |||||
QY 1303 CCCAGCCACAAAGGGCGCGCGCGCAACTTCTGCAAGAACCGCAGCGAGCGCGCGCGCC 1362
    |||||
Db 440 ProSerH1sLysGlyArGProGlyAsnPhelLeuGlnAsnArGProGlu----- 455
    |||||
QY 1363 ACCGTGCCCCACCGCCCCCGCGGAGAGCTTCCGCTTC---GAGGAGACCAACCCCGCC 1419
    |||||
Db 456 -----ProThrAlaProProAlaGlnSerPhelGlyGlnGlu11eThrProSer 473
    |||||
QY 1420 CCCAAGCAAGAGCCCAAGAGACCGGAGCCCTACCGGAGCCCTGACCGCCCTGCGCACG 1479
    |||||
Db 474 GlnLysGlnGlnGlnLysAsnGlnGluPro-----GlnProLeuThrSerLeuArGser 491
    |||||
QY 1480 CTGTTGGGAGCGCGCCCCCTG 1500
    |||||
Db 492 LeuPhelGlyAsnAspProLeu 498
    |||||
```

Search completed: March 11, 2005, 16:38:47  
Job time : 210.272 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:05:25 ; Search time 42.6742 Seconds  
(without alignments)  
5279.323 Million cell updates/sec

Title: US-09-475-704A-4  
Perfect score: 2840  
Sequence: 1 atggcgccgcgcgcagcat.....ggcgccctctgacagtaa 1509

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1 513545 seqs, 74649064 residues  
Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ndp.model -DEV=x1p  
-Q/cgn2\_1/USPTO.epool.p/US09475704/runat\_10032005\_140222\_14851/app\_query.fasta\_1.3342  
-DB=Issued Patents\_AA -QFMT=fastaan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=ICL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704.CCEN\_1.1.86 @runat\_10032005\_140222\_14851 -NCPu=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents\_AA.\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/1aa/pctus\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	86.5	492	4	US-09-991-258-5
2	2246	79.1	500	1	US-08-375-510-1
3	2246	79.1	500	2	US-08-487-657-1
4	2246	79.1	500	4	US-09-309-572-16
5	2246	79.1	500	4	US-09-718-096-16
6	2238	78.8	1350	4	US-09-952-060-35
7	2237	78.8	512	3	US-08-463-210-8
8	2237	78.8	512	3	US-09-124-900-2
9	2237	78.8	512	4	US-08-463-028-8
10	2223	78.3	500	2	US-08-816-1558-45
11	2223	78.3	500	3	US-09-079-587-45
12	2094	73.7	493	4	US-09-952-060-33

13	2018.5	71.1	478	1	US-08-127-499A-11	Sequence 11, Appl
14	2018.5	70.8	437	1	US-08-482-847-11	Sequence 11, Appl
15	2011	70.8	437	3	US-08-392-794A-2	Sequence 2, Appl1
16	1949	68.6	512	4	US-09-319-588C-4	Sequence 4, Appl1
17	1748.5	61.6	498	1	US-08-470-202-59	Sequence 59, Appl
18	1748.5	61.6	498	1	US-08-471-702-59	Sequence 59, Appl
19	1748.5	61.6	498	2	US-08-468-059-59	Sequence 59, Appl
20	1748.5	61.6	498	3	US-09-109-916-59	Sequence 59, Appl
21	1748.5	61.6	498	4	US-09-886-156-59	Sequence 59, Appl
22	1748.5	61.6	498	4	US-09-886-149-59	Sequence 59, Appl
23	1748.5	61.6	498	4	US-09-886-150-59	Sequence 59, Appl
24	1748.5	61.6	498	4	US-09-886-159-59	Sequence 59, Appl
25	1748.5	61.6	498	4	US-10-326-090-59	Sequence 59, Appl
26	1726.5	60.8	498	1	US-08-470-202-60	Sequence 60, Appl
27	1726.5	60.8	498	1	US-08-471-702-60	Sequence 60, Appl
28	1726.5	60.8	498	2	US-08-468-059-60	Sequence 60, Appl
29	1726.5	60.8	498	3	US-09-109-916-60	Sequence 60, Appl
30	1726.5	60.8	498	4	US-09-886-156-60	Sequence 60, Appl
31	1726.5	60.8	498	4	US-09-886-149-60	Sequence 60, Appl
32	1726.5	60.8	498	4	US-09-886-150-60	Sequence 60, Appl
33	1726.5	60.8	498	4	US-09-886-159-60	Sequence 60, Appl
34	1726.5	60.8	498	4	US-10-326-090-60	Sequence 60, Appl
35	1671.5	58.9	363	2	US-08-850-049-130	Sequence 130, App
36	1671.5	58.9	363	2	US-08-050-478-130	Sequence 130, App
37	1671.5	58.9	363	3	US-09-414-117-130	Sequence 130, App
38	1671.5	58.9	363	4	US-09-678-437-130	Sequence 130, App
39	1671.5	58.9	363	4	US-09-943-722-130	Sequence 130, App
40	1545.5	54.4	799	1	US-07-648-796A-7	Sequence 7, Appl1
41	1544	54.4	458	1	US-07-648-796A-5	Sequence 5, Appl1
42	1530.5	53.3	337	1	US-07-648-796A-1	Sequence 1, Appl1
43	1514.5	53.3	328	4	US-08-776-188C-77	Sequence 77, Appl
44	1453.5	51.2	518	4	US-09-206-551-45	Sequence 45, Appl
45	1410.5	49.7	294	3	US-09-370-368-3	Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-991-258-5  
; Sequence 5, Application US/09991258  
; Patent No. 6783939  
; GENERAL INFORMATION:  
; APPLICANT: Olmsted, Robert  
; APPLICANT: Keith, Paula  
; APPLICANT: Dryga, Sergey  
; APPLICANT: Maughan, Maureen  
; APPLICANT: Johnston, Robert  
; APPLICANT: Davis, Nancy  
; APPLICANT: Swannstrom, Ronald  
; TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSONES WITH MODIFIED HIV GENES FOR USE  
; FILE REFERENCE: 0113.0001U3  
; CURRENT APPLICATION NUMBER: US/09/991,258  
; CURRENT FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: 09/902,537  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; No. 6783939 =  
US-09-991-258-5  
Alignment Scores: 2,32e+141 Length: 492  
Pred. No.: 2457.00 Matches: 465  
Score:





```
TELEFAX: (703) 241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Human immunodeficiency virus type 1
US-08-375-510-1

Alignment Scores:
Pred. No.: 1,42e-128 Length: 500
Score: 2246.00 Matches: 425
Percent Similarity: 90.16% Conservative: 33
Best Local Similarity: 83.66% Mismatches: 36
Query Match: 79.08% Indels: 14
DB: 1 Gaps: 5

US-09-475-704a-4 (1-1509) x US-08-375-510-1 (1-500)

QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGAGAGAGAGAGAGAGAGATCCGC 60
DB 1 MetGlyAlaArgLaserValLeuSerGlyGlyLeuAspArgTrpGlySerIleArg 20
QY 61 CTGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 21 LeuArgProGlyGlySerGlySerGlySerGlySerGlySerGlySerGly 40
QY 121 CTGGAGAGGCTTGGCTTGAACCCGCGCTGCTGAGAGAGAGAGAGAGAGATC 180
DB 41 LeuGluArgPheLeuValLeuProGlyLeuLeuGluThrSerGlyCysArgGlnIle 60
QY 181 ATGAAGCACTGAGAGCGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGATC 240
DB 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlyGlnLeuArgSerLeuTyrAsn 80
QY 241 ACCGAGCGCGCGCGCGCATCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAG 300
DB 81 ThrIleAlaValLeuTyrCysValIleGlnArgIleAspValIleAspThrIleGlnIle 100
QY 301 CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 101 LeuAspIleIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 120
QY 361 GCCGAGCGC-----AAGTGAAGCGAGAGAGAGAGAGAGAGAGAGAGAG 411
DB 121 AspThrGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 140
QY 412 CAGATGGTGCACAGAGCGCATGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAG 471
DB 141 GlnMetValIleGlnIleIleSerProGlnThrLeuAsnAlaTrpAlaIleValIleGln 160
QY 472 GAGAGAGCGCTTGAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
DB 161 GluIleValIlePheSerProGlnValIleProMetPheSerAlaLeuSerGlnIleAlaThr 180
QY 532 CCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 591
DB 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyIleGlnIleAlaIleMetGlnMet 200
QY 592 CTGAAGAGACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651
DB 201 LeuArgGlnThrIleAsnGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 220
QY 652 GGGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 711
DB 221 GlyProIleAlaIleProGlyGlnMetArgGlnProArgGlySerAspIleIleGlyThrThr 240
```

```
QY 712 AGCACCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771
DB 241 SerThrLeuGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 260
QY 772 ATCTCAAGCGGTGATCATCTGAGCGCTGAGAGAGAGAGAGAGAGAGAGAGAG 831
DB 261 IleTyrIleArgTrpIleIleLeuGlyLeuAsnIleIleValArgMetIleSerProThr 280
QY 832 AGCATCTGAGACATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891
DB 281 SerIleLeuAspIleArgGlnIleGlyProGlySerProPheArgAspTyrValAspArgPhe 300
QY 892 TTCAGAGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951
DB 301 TyrIleThrLeuArgIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 320
QY 952 CTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011
DB 321 LeuLeuValGlnAsnIleAsnProAspCysIleThrIleLeuValAlaLeuGlyProGly 340
QY 1012 GCCACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071
DB 341 AlaThrLeuGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 360
QY 1072 CGCGTGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128
DB 361 ArgValAlaLeuAlaGlnIleMetSerGlnValThrAsnProAlaThrIleMetIleGlnIle 380
QY 1129 AGCAACTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188
DB 381 GlyAsnPheArgAsnGlnIleArgIleThrValIleCysPheAsnIleGlySerGlnIle 400
QY 1189 ATGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1248
DB 401 IleAlaIleAsnIleCysArgIleProArgIleGlySerGlySerGlySerGlySerGly 420
QY 1249 CACGAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308
DB 421 HisGlnMetLeuAspCysThrGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 440
QY 1309 CACAAGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368
DB 441 HisIleArgIleArgProGlyIleAsnIleGlnIleGlnIleGlnIleGlnIleGlnIle 454
QY 1369 CCCACGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1422
DB 455 ProThrAlaProProGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 474
QY 1423 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482
DB 475 LysGlnIleProIleAspIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 492
QY 1483 TTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
DB 493 PheGlySerAspProSerSerGln 500

RESULT 3
US-08-487-657-1
Sequence 1, Application US/08487657
Patent No. 5834267
GENERAL INFORMATION:
APPLICANT: Saito, Atsushi
APPLICANT: Sinagawa, Hideo
APPLICANT: Nakata, Atsuo
TITLE OF INVENTION: HIV ANTIGEN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22040-0747
```

```

1 COMPUTER READABLE FORM:
2 MEDIUM TYPE: floppy disk
3 COMPUTER: IBM PC compatible
4 OPERATING SYSTEM: PC-DOS/MS-DOS
5 SOFTWARE: Patentin Release #1.0, Version #1.2
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/08/487,657
8 FILING DATE: 07-JUN-1995
9 CLASSIFICATION: 435
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: US/08/375,510
12 FILING DATE: 18-JUN-1995
13 APPLICATION NUMBER: US/07/985,949
14 FILING DATE: 04-DEC-1992
15 ATTORNEY/AGENT INFORMATION:
16 NAME: Svensson, Leonard R.
17 REGISTRATION NUMBER: 30,330
18 REFERENCE/DOCKET NUMBER: 216-309P
19 TELECOMMUNICATION INFORMATION:
20 TELEPHONE: (703) 241-1300
21 TELEFAX: (703) 241-2848
22 TELEX: 248345
23 INFORMATION FOR SEQ ID NO: 1:
24 SEQUENCE CHARACTERISTICS:
25 LENGTH: 500 amino acids
26 TYPE: amino acid
27 STRANDEDNESS: single
28 TOPOLOGY: linear
29 MOLECULE TYPE: protein
30 HYPOTHEetical: NO
31 ANTI-SENSE: NO
32 ORIGINAL SOURCE:
33 ORGANISM: Human immunodeficiency virus type 1
34
35 US-08-487-657-1

```

```
Db      493 PheGlySerAspProSerSerGln 500
RESULT 4
US-09-309-572-16
; Sequence 16, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; EARLIER FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: gag polyprotein
US-09-309-572-16

Alignment Scores:
Pred. No.:      1,42e-128      Length:      500
Score:          2246.00        Matches:     425
Percent Similarity: 90.16%      Conservative: 33
Best Local Similarity: 83.66%    Mismatches:  36
Query Match:    79.08%         Indels:      14
DB:             4             Gaps:        5

US-09-475-704A-4 (1-1509) x US-09-309-572-16 (1-500)
QY      1 ATGGGCGCGCGCGCGAGCATCTCCGCGCGAGAGAGGAGCAAGTGGGAGGAATCCGC 60
DB      1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspLysTrpGluLysIleArg 20
QY      61 CTGGCGCGCGCGCGGAGAGAGCACTACATGCTGAGACCACTGGTGTGGCCAGCCGCGAG 120
DB      21 LeuArgProGlyGlyLysLysGlnTyrLysLeuLysHisIleValTrpAlaSerArgGlu 40
QY      121 CTGGAGGGCTTGCGCTGAAACCCCGCGCTGCTGGAACCGCGAGGGCTGGCAAGCATG 180
DB      41 LeuGluArgPheAlaValAsnProGlyLeuLeuGlnThrSerGlnLysCysArgGlnIle 60
QY      181 ATGAAGCAGCTGAGCCGCGCGCTGAGACCGGACCGAGAGGCTGGCGAGCGCTGTACAC 240
DB      61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGlnLeuArgSerLeuTyrAsn 80
QY      241 ACCGTGGCCACCTGTACTGCTGTGACGCGCGCGCATGAGGTCCGACACCAAGAGGCC 300
DB      81 ThrIleAlaValLeuTyrCysValHisGlnArgIleAspValLysAspThrLysGlnAla 100
QY      301 CTGGACAAATGAGAGAGAGAGCAAGTCCCAAGACAAAGAACCAAGCGCGCAAGAG 360
DB      101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
QY      361 GCCGAGCGG-----AAGGTAGCCAGAACTACCCCATCGTGTGAGAACCTGACGGGC 411
DB      121 AspThrGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 140
QY      412 CAGATGTGTACACAGGCGCATCAGCCCGCGACCCCTGTAACGCTGGGTGAAGTGAATGAG 471
DB      141 GlnMetValHisGlnIleHisSerProArgThrLeuAsnAlaIleTrpValLysValGln 160
QY      472 GAGAGGCGCTTACAGCCCGCAGAGTATCCCATGTTACACCGCTGTAGCGAGCGCGCAC 531
DB      161 GlnLysAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGlnGlyAlaThr 180
QY      532 CCCGAGCCTGAACAAGATGTTGAACACCGTGGCGCGCACAGGCGCGCATGACAGT 591
```

```
DB      181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnIleAlaMetGlnMet 200
QY      592 CTGAAGACACCATCAACAGAGAGCGCGCGAGTGGGACCGCTGCACCCCGTGCAGGCC 651
DB      201 LeuLysGlnThrIleAsnGlnGlnAlaIleGlnTrpAspArgLeuHisProValHisAla 220
QY      652 GGGCGCGTGGCGCGCGCGAGATGGCGCAACCCCGCGCGAGAGCAATCCGCGCGCGCAC 711
DB      221 GlyProIleAlaProGlyGlnMetArgGlnProArgGlySerAspIleAlaGlyThrThr 240
QY      712 AGCACCCTGACAGAGAGAGATCCGCTGATGACAGACCAACCCCGCTGGCGCGTGGAGCAG 771
DB      241 SerThrLeuGlnGlnGlnIleGlyTyrMetThrHisAsnProIleProValGlyGln 260
QY      772 ATCTACAGCGGTGATCATCTGCGCTGAAACAAGATGTGGATGTACAGCCCGCTG 831
DB      261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThr 280
QY      832 AGCATCTGTGACATCCGCGAGGCGCGCGAGGCGCTTCCGCGACATGACGCGCTTC 891
DB      281 SerIleLeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPhe 300
QY      892 TTCAGAGCCTGCGCGCGCGAGAGCGCACCCAGACGTGAAGATCTGATGACCGAGCC 951
DB      301 TyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTrpMetThrGlnThr 320
QY      952 CTGCTGTGTGACAGACCGCAACCCCGCATGCAAGACCATCTGCGCGCTTCCGCGCGCG 1011
DB      321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGly 340
QY      1012 GCCACCCCTGAGAGAGATGATGACCGCGCTGCGAGGCGTGGCGCGCGCGCGCAAGAGCC 1071
DB      341 AlaThrLeuGlnGlnMetMetThrLysCysGlnGlyValGlyLysProGlnLysLysAla 360
QY      1072 CCGCTGTGTGCGCGCGAGCGATGAGCGCAG---GCCAACAGCGTAAATCATGATGACGAAG 1128
DB      361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLys 380
QY      1129 AGCAACTTCAAGGCGCGCGCGCAACGTCAAGTGTCACTGCGCGCAAGAGGCGCAC 1188
DB      381 GlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGlnGlyHis 400
QY      1189 ATCGCCAGAAATCTGCGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGGCGCGAGAGGCC 1248
DB      401 IleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpLysGlyGlyGlnGly 420
QY      1249 CACGAGATGAAGATGTCACCGAGCGCGCGCAACTTCTGAGGCAAGATCTGGCGCGCAGC 1308
DB      421 HisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440
QY      1309 CACAAGGCGCGCGCGCGCAACTTCTGAGAAACCGACGAGACCGCGCGCGCACCGTGC 1368
DB      441 HisLysGlyArgProGlyLysAsnPheLeuGlnSerArgProGlu----- 454
QY      1369 CCCACCGCGCGCGCGCGCGAGGCTTCCGCTTC-----GAGAGACCAACCCCGCGCGCC 1422
DB      455 ProThrAlaProProGlnGlnSerPheArgPheGlyGlnGlnThrThrProSerGln 474
QY      1423 AAGCAGAGCCCAAGAGACCGAGCGCTACCGCGAGCGCGCTGACCGCGCTGGCGAGCTG 1482
DB      475 LysGlnGlnProIleAspLysGlnLeuTyr-----ProLeuAlaSerLeuArgSerLeu 492
QY      1483 TTGCGCAGCGCGCGCGCTGAGCCAG 1506
DB      493 PheGlySerAspProSerSerGln 500

RESULT 5
US-09-718-096-16
; Sequence 16, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Melke-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
```

```
/ FILE REFERENCE: 35-195
/ CURRENT APPLICATION NUMBER: US/09/718, 096
/ PRIOR FILING DATE: 2000-11-22
/ PRIOR APPLICATION NUMBER: DE 19856463.5
/ PRIOR FILING DATE: 1998-11-26
/ PRIOR APPLICATION NUMBER: EP 99250415.9
/ PRIOR FILING DATE: 1999-11-25
/ PRIOR APPLICATION NUMBER: US 09/309, 572
/ PRIOR FILING DATE: 1999-05-11
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 16
/ LENGTH: 500
/ TYPE: PRT
/ ORGANISM: Human immunodeficiency virus type 1
/ FEATURE:
/ OTHER INFORMATION: gag polyprotein
/ US-09-718-096-16
```

## Alignment Scores:

Pred. No.:	1,42e-128	Length:	500
Score:	2246.00	Matches:	425
Percent Similarity:	90.16%	Conservative:	33
Best Local Similarity:	83.66%	Mismatches:	36
Query Match:	79.08%	Indels:	14
DB:	4	Gaps:	5

US-09-475-704a-4 (1-1509) x US-09-718-096-16 (1-500)

```
QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGAGAGCTGGAGAGATCCGC 60
DB 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspArgTrpGluLysIleArg 20
QY 61 CTGGCGCGCGCGCGCGAGAAAGCACTGATGCTGAAGCACTGCTGGCGCGCGAG 120
DB 21 LeuArgProGlyGlyLysLysGlnTyrLysLeuYsnIleValIleTrpAlaSerArgGlu 40
QY 121 CTGGAAGGCTTGGCGCGTGAACCGCGCTGTGGAACCGCGAGGGCTGCAAGAGATC 180
DB 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGlnThrSerGlnLysArgGlnIle 60
QY 181 ATGAAGAGCTGCAAGCGCGCGCTGCAAGCGCGAGAGCTGGAGCTGTATCAAC 240
DB 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGlnLeuArgSerLeuTyrAsn 80
QY 241 ACCGCGCGCGCGCTGATCTGCTGCAAGCGCGCGATGAGTCCGCGACCAAGAGGCG 300
DB 81 ThrIleAlaValLeuTyrCysValIleValArgGlyLeuAspValLysAspThrLysGlnIle 100
QY 301 CTGGAAGAGCTGCAAGCGCGCGCTGCAAGCGCGAGAGCTGGAGCTGTATCAAC 360
DB 101 LeuArgLysIleLeuGlnGlnLysLeuYsnLysLeuValLeuGlnIleAlaIle 120
QY 361 GCCGACGCG-----AAGTGAAGCAAACTACCCCATGTGTGCAAGCTGCAAGGCG 411
DB 121 AspThrGlyAsnAsnSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGlnIly 140
QY 412 CAGATGCTGCAAGCGCGCGCTGCAAGCGCGAGAGCTGGAGTGAAGTGAATCCAG 471
DB 141 GlnMetValIleValGlnIleIleSerProArgIleLeuAsnIleTrpAlaLysValIleGlu 160
QY 472 GAGAAAGCTTCAAGCGCGCGAGTGAATCCCATGTTCACCGCGCTGAGCGAGCGCGCAC 531
DB 161 GlnLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGlnLysIleThr 180
QY 532 CCCCGAGCTTGAACAGATGTTTAAACACCGTGGCGCGCGCACGAGCGCGCATGAGATG 591
DB 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnIleGlnAlaIleMetGlnMet 200
QY 592 CTGAAGAGCACCATGAAGAGAGCGCGCGAGTGGAGCGCGCTGACACCGCGCGAGCGC 651
DB 201 LeuLysGlnThrIleAsnGlnIleAlaIleGlnIleTrpAspArgLeuIleProValIleAla 220
```

```
QY 652 GGCGCGCGCGCGCGCGCGAGATGGCGGACCGCGCGGAGGAGATGCGCGCGCGCAC 711
DB 221 GlyProIleAlaProGlyIleMetArgIleProAspGlySerAspIleAlaGlyThrThr 240
QY 712 AGCACCGCTGAGAGAGAGATGCGCTGATGACCAAGACCGCGCGCGCTGGAGCGAG 771
DB 241 SerThrLeuGlnGlnIleGlyTyrMetThrIleAsnProIleProIleProIleGlyGlu 260
QY 772 ATCTAAGACCGGTGATCATCTGTGGCTGAAACAAGATGCGAGATGTACACCGCGGTG 831
DB 261 IleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProThr 280
QY 832 AGCATCTGGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 891
DB 281 SerIleLeuAspIleArgGlnIleProLysGlnProPheArgAspTyrValAspArgPhe 300
QY 892 TTCAAGACCGCTGCGCGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951
DB 301 TyrLysThrLeuArgAlaGlnIleAlaSerGlnIleValLysAsnThrMetThrGluThr 320
QY 952 CTGCTGCTGCGAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1011
DB 321 LeuLeuValIleGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGly 340
QY 1012 GCCACCGTGAAGAGATGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1071
DB 341 AlaThrLeuGlnGlnMetMetThrAlaCysGlnIleValGlyGlyProGlyIleLysAla 360
QY 1072 CGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1128
DB 361 ArgValIleAlaGlnAlaIleMetSerGlnValThrAsnProAlaThrIleMetIleGlnLys 380
QY 1129 AGCAACTTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1188
DB 381 GlyAsnPheArgAsnGlnIleArgLysThrValLysCysPheAsnGlyLysGlnIleLys 400
QY 1189 ATGCGCAAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1248
DB 401 IleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnIle 420
QY 1249 CACCGATGAAGAGCTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1308
DB 421 HisGlnMetLysAspCysThrGlnArgGlnIleAsnPheLeuGlyLysIleTrpProSer 440
QY 1309 CACAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1368
DB 441 HisLysGlyArgProGlyAsnPheLeuGlnSerArgProGlu----- 454
QY 1369 CCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1422
DB 455 ProThrAlaProProGlnGlnLysSerPheArgPheGlyGlnGluThrThrThrProSerGln 474
QY 1423 AAGCAGAGGCCCAAGAGACCGGAGCGCTTACCGCGAGCGCGCTGACCGCGCGCGCTG 1482
DB 475 LysGlnGlnProIleAspLysGlnLeuTyr-----ProLeuAlaSerLeuArgSerLeu 492
QY 1483 TTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1506
DB 493 PheGlySerAspProSerSerGln 500
```

RESULT 6  
US-09-952-060-35  
Sequence 35, Application US/09952060  
Patent No. 6733993  
GENERAL INFORMATION:  
APPLICANT: Emili, Emilio A.  
APPLICANT: Youil, Rima  
APPLICANT: Bett, Andrew J.  
APPLICANT: Chen, Ling  
APPLICANT: Kaelow, David C.  
APPLICANT: Shiver, John W.  
APPLICANT: Toner, Timothy J.  
APPLICANT: Casimiro, Danilo R.

```
/ TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
/ TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV-1 GAG, POL, NEF AND
/ TITLE OF INVENTION: MODIFICATIONS
/ FILE REFERENCE: 20747Y
/ CURRENT APPLICATION NUMBER: US/09/952,060
/ CURRENT FILING DATE: 2001-09-14
/ PRIOR APPLICATION NUMBER: PCT/US01/28861
/ PRIOR FILING DATE: 2001-09-14
/ PRIOR APPLICATION NUMBER: 60/317,814
/ PRIOR FILING DATE: 2001-09-07
/ PRIOR APPLICATION NUMBER: 60/279,056
/ PRIOR FILING DATE: 2001-03-27
/ PRIOR APPLICATION NUMBER: 60/233,180
/ PRIOR FILING DATE: 2000-09-15
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 35
/ LENGTH: 1350
/ TYPE: PRF
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Codon optimized gag-1A pol fusion
/ US-09-952-060-35

Alignment Scores:
Pred. No.: 5.01e-128 Length: 1350
Score: 2238.00 Matches: 423
Percent Similarity: 90.35% Conservative: 36
Best Local Similarity: 83.27% Mismatches: 35
Query Match: 78.80% Indels: 14
DB: 4 Gaps: 5

US-09-475-704a-4 (1-1509) x US-09-952-060-35 (1-1350)
QY 1 ATGGGCGCCGCGCCGACATCTGCGCGGAGAGAGTGGACAACTGGAGAGATCCGC 60
DB 1 MetGlyAlaArgHisLeuSerGlyGluLeuHisPheTyrPheValGluGlnGly 20
QY 61 CTGGCGCCGCGCGGCAAGAACTACATGCTTGAAGACCTGTGTGGCGCCGCGGAG 120
DB 21 LeuArgProGlyGlyLeuValSerGlyLeuValSerGlyLeuValSerGlyLeu 40
QY 121 CTGGAGGCGCTGCGCGGCAAGCGCGCGCTGGAAGCGCGCGCGCGCGCGCGAGTC 180
DB 41 LeuGluArgPheAlaValAlaPheProGlyLeuLeuGlnHisSerGlyCysArgGln 60
QY 181 ATGAGCAGCTGAGCGCGCGCTGAGACCGCGCGCGAGGAGTGGCGAGCTGTACAC 240
DB 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnLeuHisSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGCTGCGACGCGCGCATGAGGTCCGCAACCAAGAGGCC 300
DB 81 ThrValAlaThrLeuTyrCysValHisGlnLeuValSerValHisPheThrVal 100
QY 301 CTGGAGAAATGAGAGAGAGAGAGAGAGTCCACAGAGAGAGAGAGAGAGAGAGAG 360
DB 101 LeuGlnValSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
QY 361 GCCGAGCGC-----AAGGTGAGCAGAACTACCCCTGCTGGAAGCTGAGAGC 411
DB 121 GlyThrGlyAsnSerSerGlnValSerGlnHisPheProLeuValGlnHisGlnGln 140
QY 412 CAGATGTCACACGAGCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 471
DB 141 GlnMetValHisGlnAlaHisSerProArgThrLeuHisPheValValValGln 160
QY 472 GAGAGGCGCTTGAAGCGCGCGAGAGTCCCATTTGACCGCGCGCGCGCGCGCGCG 531
DB 161 GlnValAlaPheSerProGlnValHisPheMetPheSerLeuLeuSerGlnGlnAla 180
QY 532 CCCGAGAGCTTGAAGAGTGTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGAGT 591
DB 181 ProGlnAsnLeuHisPheThrLeuHisPheValGlyGlnHisGlnAlaHisMetGln 200
```

```
QY 592 CTGAAGAGACCATCAACGAGAGCGCGCGAGTGGAGACCGCGCGCGCGCGCGCG 651
DB 201 LeuValGlnThrLeuHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220
QY 652 GCGCGCGCGCGCGCGCGAGATGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
DB 221 GlyProIleAlaProIleGlnMetCysGlnProArgGlySerHisPheAlaGlyThr 240
QY 712 AGCAGCTGAGAGAGAGATCGCTGATGACGAGACCGCGCGCGCGCGCGCGCGCG 771
DB 241 SerThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
QY 772 ATCTAAGCGGTGATCACTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 831
DB 261 IleTyrLeuArgTyrIleLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280
QY 832 AGCATCTGAGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 891
DB 281 SerIleLeuHisPheLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
QY 892 TTCAGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951
DB 301 TyrLeuThrLeuArgHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320
QY 952 CTGCTGTGTCAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1011
DB 321 LeuLeuValGlnHisPheAlaHisPheProArgCysValThrIleLeuHisPheLeu 340
QY 1012 GCCACCTGAGAGAGATGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1071
DB 341 AlaThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 360
QY 1072 CCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1128
DB 361 ArgValLeuAlaGlnHisPheSerGlnValThrAsnSerAlaThrIleMetCysArg 380
QY 1129 AGCACTTCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1188
DB 381 GlyAsnPheArgHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 400
QY 1189 ATGCGCAAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1248
DB 401 IleAlaValHisCysArgAlaProArgGlySerGlyCysThrPheCysGlySerGln 420
QY 1249 CACGAGTAAAGAGCTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1308
DB 421 HisGlnMetLeuAspCysHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 440
QY 1309 CACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1368
DB 441 HisValGlyArgProIleHisPheLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 454
QY 1369 CCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1422
DB 455 ProThrAlaProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 474
QY 1423 AAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1482
DB 475 ValGlnGlnProIleHisPheSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 492
QY 1483 TTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1506
DB 493 PheGlyAsnAspProSerSerGln 500

RESULT 7
US-08-463-210-8
/ Sequence 8, Application US/08463210
/ Patent No. 6001977
/ GENERAL INFORMATION:
/ APPLICANT: CHANG, Nancy T.
/ APPLICANT: GALLO, Robert C.
/ APPLICANT: WONG-STALL, Ploesie
```

```

/ TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Morgan & Finnegan, L.L.P.
/ STREET: 345 Park Avenue
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10154-0053
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: IBM PC compatible
/ SOFTWARE: Patent Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US 06/463,210
/ FILING DATE: 05-JUN-1995
/ CLASSIFICATION: 436
/ PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: US 06/693,866
/ FILING DATE: 23-JAN-1985
/ PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: US 06/659,339
/ FILING DATE: 10-OCT-1984
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Serunian, Leslie A.
/ REGISTRATION NUMBER: 35,353
/ REFERENCE/DOCKET NUMBER: 2026-4193US2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 758-4800
/ TELEFAX: (212) 751-6849
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 512 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: NO
/ ORIGINAL SOURCE:
/ ORGANISM: HTLV-III
/ FEATURE:
/ NAME/KEY: Protein
/ LOCATION: 1..512
/ OTHER INFORMATION: /note="gag protein of HTLV-III"
/
US-08-463-210-8
/
Alignment Scores:
Pred. No.: 4,99e-128 Length: 512
Score: 2237.00 Matches: 422
Percent Similarity: 89.69% Conservative: 39
Best Local Similarity: 82.10% Mismatches: 39
Query Match: 78.77% Indels: 14
DB: 3 Gaps: 5
/
US-09-475-704A-4 (1-1509) x US-08-463-210-8 (1-512)
QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGCGAGAGCTGACAGTGGAGAGATCCG 60
DB 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTTrpGlnLysIleArg 20
QY 61 CTGCGCGCGCGCGCGAGAGAGCACTACATGCTGAGACCTGAGTGGCGCGAGCGCGAG 120
DB 21 LeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTTrpAlaSerArgGln 40
QY 121 CTGAGAGGCTTGCGCGCTGAACCCCGCGCTGCTGAGAGACCGCGAGGCTGCAAGAGATC 180
DB 41 LeuGluArgPheAlaValAlaAsnProGlyLeuLeuGluThrSerGluGlyCysArgGlnIle 60
QY 181 ATGAAGACGCTGACGCGCGCGCTGAGACCGCGAGACCGGACCTGCGAGCTGTATACAC 240
DB 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGluLeuArgSerLeuTyrAsn 80

```

```

QY 241 ACCGTGGCCACCTGTACTGCTGACAGCGCGGATCGAGTCCGCGACACCAAGAGGCC 300
DB 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluIleLysAspThrLysGluAla 100
QY 301 CTGACACAGATCGAGAGAGAGAGAGAGCAAGTCCCGACAGAGAGACCGAGGCGCAAGAG 360
DB 101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
QY 361 GCCGACGCGC-----AAGGTGAGCCAGAACTACCCCATGTCGACAACTGCAAGGCG 411
DB 121 AspThrGlnLysSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIleGlnGly 140
QY 412 CAGATGTGACACAGGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 471
DB 141 GlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTTrpValLysValGlu 160
QY 472 GAGAAAGCGCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 531
DB 161 GlyLysAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGlnValAlaThr 180
QY 532 CCCGAGACCTGAACAGATGTTGAACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCATG 591
DB 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMet 200
QY 592 CTGAAGGACACCATCAAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651
DB 201 LeuLysGluThrIleAsnGlnGlnAlaIleGluTTrpAspArgValHisProValHisAla 220
QY 652 GCGCGCGCGCGCGCGCGCGCGAGATGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
DB 221 GlyProIleAlaProGlyGlnMetArgGlnProArgGlySerAspIleAlaGlyThrThr 240
QY 712 AGCACCTGCGAGAGACAGATGCGCTGATGACACGACACCGCGCGCGCGCGCGCGCGCG 771
DB 241 SerThrLeuGlnGlnGlnIleGlyTyrMetThrAsnAsnProProIleProValGlyGlu 260
QY 772 ATCTCAAGCGGTGATCATCTGCGCGCTGAAACAGATCGTGGCATGACAGCGCGCGT 831
DB 261 IleTyrLysArgTTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProThr 280
QY 832 AGCATCTGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 891
DB 281 SerIleLeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPhe 300
QY 892 TTCAAGACCTTCGCGCGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951
DB 301 TyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTTrpMetThrGlnThr 320
QY 952 CTGCTGTGTCAGAAAGCGCAACCGCGCATGCAAGACATCTGCGCGCTCGCGCGCGCGCG 1011
DB 321 LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340
QY 1012 GCCACCTGAGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCGCGCGCGCGCGCGCG 1071
DB 341 AlaThrLeuGlnGlnGlnMetCThrAlaCysGlnGlyValGlyGlyProGlyHisLysAla 360
QY 1072 CGCGTGTGCGCGCGCGCGCATAGCCAG---GCCAACACCGTGAACATATATGAGAGAG 1128
DB 361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGlnArg 380
QY 1129 AGCAACTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1188
DB 381 GlyAsnPheAspGlnArgLysLysMetValLysCysPheAsnCysGlyLysGlnGlyHis 400
QY 1189 ATGCGCAAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1248
DB 401 ThrAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGly 420
QY 1249 CACCAAGATGAGAGCTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1308
DB 421 HisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLysGlyLysIleTrpProSer 440
QY 1309 CACAAGGCGCGCGCGCGCACTTCTGAGAAACGACGAGCGCGCGCGCGCGCGCGCGCG 1368

```









```

Oy      1012 GCCACCCCTGAGGAATATGTACCGCCTTCGCACGGGCGTGGGGGCCCGGCAAGAAGCC 107
Db      341 ALeThrLeuGlumetmetThraLaCySgInsglyValGlyelYProGIYhiLysAla 360
Oy      1072 CGCGTGCTGGCCGAGCGCATGAGCCAG--GCCAACAGCGTGAMCATCATGATGCAAG 1128
Db      361 ArgValLeuAlaGluAlaMetSerGlnValThrAenThraLaThrilemetetGlnArg 380
Oy      1129 AGCAACTTCMAAGGCCCCCGGGGCAAGCTCAAGTCTTCAACTGCGGCAAGAGGCCAC 1186
Db      381 GlyAenPheArgAenGlnhArgIylsmetVallyScySPheaSCySGilyLySgluGIYhiS 400
Oy      1189 ATCCGACAAGAACTGCGCGCGCCCCCGGCAAGAAAGGCGTGTGGAACTGGGGCAAGAGGGC 1248
Db      401 ThrAlaArgAenCySArghAlaProArgIylslySGlyCYSTripLYscySGilyLySgluGIY 420
Oy      1249 CACCAGATGAAGACTGTGACCGAGCGCCAGGCACTTCCTGGCGMAATGTGCGCCAGC 1308
Db      421 HieGlnmetLyYabpCYstHrglunArgGlnAlaAenPheleuGIYlyelIetripProber 440
Oy      1309 CACAAAGGCGCGCCCCGGGCACTTCTGTGAGAAcCGCACGAGACCGCGCGCCCCCAccGTG 1368
Db      441 TyrlySGIYAArgProGIYAbenPheleuGInserArGProGIUProThAlaProPhoPe 460
Oy      1369 -----CCACGCGCGCCCCCGCGCGAGAGTTCCG-----TTCCAG 1404
Db      461 LeuGInserArGProGIUnProThralaProProGIUGInserPheArgSerGIYvalGIUn 480
Oy      1405 GAGACCAACCCCGCGCCCGCAAGCAGAGGCCCAAGACCGCGAGCCCTTAACCGAGCGCCCTG 1464
Db      481 ThrlThlrProProGlnhlySgInclunIrolleaaplySGluLeutyr-----Proleu 498
Oy      1465 ACCGCGCTGCGCAGCTGTGTGGAGCGCGGCGCTGAGCCAG 1506
Db      499 ThrsertLeuArgSerLeuPheGlyAenApProSerSerGln 512

RESULT 10
US-08-816-155B-45
/ Sequence 45, Application US/08816155B
/ Patent No. 5990091
/ GENERAL INFORMATION:
/ APPLICANT: TARTAGLIA, JAMES
/ APPLICANT: COX, WILLIAM I.
/ APPLICANT: GETTING, RUSSELL R.
/ APPLICANT: MARTINEZ, HECTOR
/ APPLICANT: PAOLETTI, ENZO
/ APPLICANT: PINCUS, STEVEN E.
/ TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
/ TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
/ NUMBER OF SEQUENCES: 48
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: FROMMER LAWRENCE & HAUG LLP
/ STREET: 745 FIFTH AVENUE
/ CITY: NEW YORK
/ STATE: NEW YORK
/ COUNTRY: USA
/ ZIP: 10151
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/Ms-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/816.155B
/ FILING DATE: 12-MAR-1997
/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: KOMALSKI, THOMAS J.
/ REGISTRATION NUMBER: 32,147
/ REFERENCE/DOCKET NUMBER: 454310-2990
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212-588-0800
/ TELEFAX: 212-588-0500

```

INFORMATION FOR SEQ ID NO: 45: SEQUENCE CHARACTERISTICS: LENGTH: 500 amino acids TYPE: amino acid STRANDEDNESS: n/a TOPOLOGY: linear MOLECULE TYPE: amino acid US-08-816-155B-45			
Alignment Scores:			
Pred. No.:	3,51e-127	Length:	500
Score:	2223.00	Matches:	420
Percent Similarity:	90.16%	Conservative:	38
Best Local Similarity:	82.68%	Mismatches:	36
Query Match:	78.27%	Indels:	14
DB:	2	Gaps:	5
US-09-475-704A-4 (1-1509) x US-08-816-155B-45 (1-500)			
QY	1	ATGGGCGCCCGGCCAGCATCTCGCGCGCGGAGAACTGGACAAGTGGAGAAATCCGC	60
DB	1	MetGlyAlaIrgAlaSerValIleuSerGlyGlyGluLeuAspArgTrpGluSileArg	20
QY	61	CTGGCGCCCGGCGGAGAAAGACATACATGTGAAGACACTGGTGTGGCGACCGCGAG	120
DB	21	LeuArgProGlyGlyIleuValuSylTrpIleuValuSileValIrrpAlaSerArgGlu	40
QY	121	CTGGAGGCGCTTCGCGCTGAACCCCGCGCTGTGAGACCGCGAGCGGTGCAGACAGATC	180
DB	41	LeuGluArgPheAlaValIleuProGlyLeuLeuGluThrSerGluGlySerArgGlnIle	60
QY	181	ATGAAGCAGCTGCAGCGCGCGCTGCAGACCGGCACCGAGAGAGTGGCGACCTGTACAC	240
DB	61	LeuGlyGlnLeuGlnProSerLeuGlnThrArgIleSerGluGluIleuArgSerLeuTrpAsn	80
QY	241	ACCGTGGCACCTCTGTATCTGGCTGCACCGCGCATGAGATCCGGACACCAAGAGGCC	300
DB	81	ThrValAlaThrLeuTrpCysValIleuSglnArgIleGluIleuAspThrLysGluAla	100
QY	301	CTGGACAAGATCAGAGAGAGACAGAAACAATGCCAGCAAGACCCAGCAGGCCAAGAG	360
DB	101	LeuAspArgSileGluGlnGluGlnIleuAsnSylSerTrpIleuValuSileGlnIleAlaIle	120
QY	361	GCCGAGGCGC-----AAGGTAGCCAGAACTACCCCATGTGTGCAGAACCTGCAGGCGC	411
DB	121	AspThrGlyIleSerAspAsnIleValSerGlnIleuTrpProIleValGlnIleuThrGlnGly	140
QY	412	CAGATGTGTGCACCAAGCCCATCAGCCCGCACCTGGAAGCGCTGGTGAAGTGAATCGAG	471
DB	141	GlnMetValIleSglnAlaIleSerProArgTrpIleuAsnAlaIrrpAlaValuValGlu	160
QY	472	GAGAAAGGCGCTCAGGCGCGGAGGTATCCCATGTTCAACGCGCTGAGCGAGGCGGCACC	531
DB	161	GluIleuValaIlePheSerProGluValIlePheMetPheSerAlaIleuSerGlnGlyAlaThr	180
QY	532	CCCCAGGACCTGAACAGATGTTGAACAACCGTGGCGCGGACACAGGCGCCGATGCAGATG	591
DB	181	ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyIleSglnIleAlaIleuMetGlnMet	200
QY	592	CTGAAGGACACCATCAACAGAGAGGCGCGCGATGGAGACCGGCTGCACCCCGTGCAGGCC	651
DB	201	LeuIleuSglnThrIleAsnGlnGluIleAlaGluTrpAspArgValIleAspValuIleAla	220
QY	652	GAGCGCGGAGCGCGCGCGCAGATGGGCGACCCCGCGAGGAGGACATCCCGCGCGCACCC	711
DB	221	GlyProIleuAlaIrrpGlyIleuMetArgGluProArgGlySerAspIleAlaGlyIleThr	240
QY	712	AGCACCCTGCAGAGACAGATCGCTGCAGATGACACCAACCCCGCGTGCACCGGCGGAC	771
DB	241	SerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnProIleProIleProValGlyGlu	260
QY	772	ATTTCAAAGCGGTGATCATCTCGGCGCTGAACAAGATGTGCGGATGTACAGCCCGGTG	831

```

Db      261  lletTlyrsArtrpillelleuenglyleuasnlyseillevaIarhmetTyrserProthr 280
QY      832  AGATCTGTGACATCCGCCAGGGCCCCAGAGCCCTTCGGCACTGACGAGCCCTTC 891
Db      281  SerilleuasnplleatrginglyProlysgluProhearghapyrValaapgaPhe 300
QY      892  TTCAGAGCCCTGGCGCCGAGAGAGCCAGCCAGTGAAGAACTGATGACCGAGAC 951
Db      301  TylTylThrleuArghaIngluInlaSerghIngluValIyAsntrpMetThrIngluThr 320
QY      952  CTGCTGTGTGAGAGAGCCCAACCCCGCACTGCAAGACCACTTCGCGCTCTCGCGCCGAG 1011
Db      321  leuIeuValglInuInlaInlaenProharpCylyThrilleuIyhalaleuuglyProhla 340
QY      1012  GCCACCTGTGAGAGAGATGATGACCGCTTCGAGGGCGTGGCGCGCCCGCCGCAAGGCC 1071
Db      341  AlarhIeuIngluInluIeuIeuThrIalCysgIngluIyValglIyProglIyIslYsAla 360
QY      1072  CGGCTGTGGCCGAGGCGATGAGCAG---GCCAACAAGCGTGAACATCAATGATGACAGAG 1128
Db      361  ArgValIeuAlaInlaIeuIeuSerghInValIThrAsnSerAlaThrilleuIeuIyAla 380
QY      1129  AGCAACTTGAAGGGCCCCCGCGCAACGTCAAGTGTCTTCAACTGCGCAAGAGAGCCAC 1188
Db      381  GlyAsnPharharghIngluIngluIyIleValIyCysePhasnCyseglIyIysgluIyHis 400
QY      1189  ATCGCCCAAGACTGCGCGCCCGCCCGCAAGAGGCTGTGTGAACTGCGGCAAGAGAGGC 1248
Db      401  ThrAlaArgAsnYsArghaIaProArgIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 420
QY      1249  CACCAATGATGAGATGACAGCGAGCGCGCAAGCCCACTTCCTGGGCAAGATGCGGCCAC 1308
Db      421  HIsghImeIyIyAspYsThrIguArgInlaIAsnPhelauIyIyIyIyIyIyIyIyIyIy 440
QY      1309  CACAAGGGCGCGCCCGCAACTTCTGCAAGAACCGGAGCGAGCGCGCCCGCCACCGTGC 1368
Db      441  TylTylsgIyArgProglIyAsnPhelauInIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 454
QY      1369  CCCACCGCGCCCGCCCGCGAGCTTCGCC---TTGAGAGAGACCAACCCCGCGCCCGCC 1422
Db      455  ProThAlaProProgluInIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 474
QY      1423  AAGCAGAGAGCCCAAGAGAGCGCGAGCCCTTACCGGAGCCCTGAGCGCGCTGGCGAGCTG 1482
Db      475  IyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 492
QY      1483  TTCGGCAGCGCGCCCGCTGAGCCAG 1506
Db      493  PheglYAsnAspProserSerghIn 500

RESULT 11
US-09-079-587-45
Sequence 45, Application US/09079587
Patent No. 613066
GENERAL INFORMATION:
APPLICANT: TARTAGLIA, JAMES
APPLICANT: COX, WILLIAM I.
APPLICANT: GETTIG, RUSSELL R.
APPLICANT: MARTINEZ, HECTOR
APPLICANT: PAOLETTI, ENZO
APPLICANT: PINCUS, STEVEN E.
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSES: FROMMER LAWRENCE & HAUG LLP
STREET: 745 FIFTH AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10151
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,587
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/816,155
FILING DATE: 12-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: KOMALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454310-2990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
STRANDEDNESS: n/a
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-09-079-587-45

Alignment Scores:
Pred. No.: 3,51e-127 Length: 500
Score: 2223.00 Matches: 420
Percent Similarity: 90.16% Conservative: 38
Best Local Similarity: 82.68% Mismatches: 36
Query Match: 78.27% Indels: 14
DB: 3 Gaps: 5

US-09-475-704A-4 (1-1509) x US-09-079-587-45 (1-500)
QY      1  ATGGCGCGCGCGCGAGCATCTGCGCGGCGAGAACTGGAACAGTGGGAGAAATCCGC 60
Db      1  MetGlyAlaArgAlaSerValIeuSerghIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 20
QY      61  CTGCGCGCGCGCGCAAGACACTACATGCTGAAGCACTGTGTGGGCCAGCGCGAG 120
Db      21  leuArghProglIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 40
QY      121  CTGAGAGGCTTGCCTGAACCCCGGCTGTGTGAAGACCGCGAGGCTGCAAGCAGATC 180
Db      41  leuIguArgPheAlaValaIAsnProglIyIeuIeuIyIyIyIyIyIyIyIyIyIyIy 60
QY      181  ATGAAGCAGCTGACGCGCCCGTGAAGACCGGAGCGGAGAGTGGCGAGCTGTATCAAC 240
Db      61  leuIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 80
QY      241  ACCGTGGCCACCTGTATCTGCTGACAGCGCGCATGAGTCCGCAACCAAGAGAGCC 300
Db      81  ThrValAlaThrIeuIyCyseValHIsghInIyIyIyIyIyIyIyIyIyIyIyIyIy 100
QY      301  CTGACCAAGATCGAGAGAGACAGAAAGTCCCAAGAGAGAGAGAGAGAGAGAGAG 360
Db      101  leuAspIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 120
QY      361  GCCGAGGCG-----AAGGTAGCGCAAGTACCCCACTGTCGAGAACCTGACAGGCG 411
Db      121  AsphTrghIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 140
QY      412  CAGATGTGACACAGGCGCATGAGCCCGGCAACCTGGAAGCGCTGGTGAAGTATGAG 471
Db      141  GlIneIyValHIsghInAlaIleSerProhArgThrIeuInlaIyIyIyIyIyIyIy 160
QY      472  GAGAAAGCTTCAAGCCCGAGAGTATCCCATGTTCACCGCGCTGAGAGAGAGAGAG 531
Db      161  GlIyIyAlaPheSerProglIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 180
QY      532  CCCGAGACCTGAACAGATGTGAACACCGTGGCGGCGCACGAGCGCGCATGCAATG 591

```

181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMet 200  
QY 592 CTGAAGACACCATCAACGAGAGGCGCGAGTGGAGCCGCTGCAACCCCGTGAAGGCC 651  
DB 201 LeuValGluThrThrLeuAsnGlnGluAlaAlaGlnThrAspArgValHisProValHisAla 220  
QY 652 GGGCGCGTGGCGCGCGCGAGTGGCGAGCCCGCGGCGGCGAGCATCGCGCGCGCGCACC 711  
DB 221 GlyProIleAlaProGlnMetArgGlnProArgGlySerAspIleAlaGlyThrThr 240  
QY 712 AGCACCCTGAGAGAGAGATCGCCTGATGACCAACCCCGCGTGGCGCGAGC 771  
DB 241 SerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnAsnProIleProValGlyGln 260  
QY 772 ATCTACAGCGGTGATCATCTCGGCGCTGAAACAAGATCGTGGATGTAACGCCCGT 831  
DB 261 IleTyrLeuArgTrpIleIleLeuGlyLeuAsnIleValArgMetTyrSerProThr 280  
QY 832 AGCATCTGAGACATCGCGCGCGCGCGAGCCCTTCGCGGACTACGTTGACCGCTTC 891  
DB 281 SerIleLeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPhe 300  
QY 892 TTCAGAGCCCTGGCGCGCGCGAGCGACCCAGCGAGCGTGAAGATGATGACCGAGCC 951  
DB 301 TyrLeuThrLeuArgIleGlnGlnAlaSerGlnGlnValIleValAsnTrpMetThrGlnThr 320  
QY 952 CTGCTGTGACAGACCCCAACCCCGACTGACAGACCATCTCGCGCGCTTCGCGCGCGC 1011  
DB 321 LeuLeuValGlnAsnAlaAsnProAspCysLeuThrIleLeuValAlaLeuGlyProAla 340  
QY 1012 GGCACCTGTGAGAGATGATGACCGCTGACAGGCGTGGCGCGCGCGCGCGCAAGGCC 1071  
DB 341 AlathThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlnHisAla 360  
QY 1072 CCGCTGCTGGCGCGCGAGCGATGAGCCAG---GCCAAGCGCTGAACATCATGATGACGAAG 1128  
DB 361 ArgValLeuAlaGlnIleMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArg 380  
QY 1129 AGCAACTTCAAGGGCGCGCGCGCAAGCTCAAGTGTCTTCAACTGCGCGCAAGAGGCCAC 1188  
DB 381 GlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnCysGlyLysGlnGlnHis 400  
QY 1189 ATCGCAACAACTGCGCGCGCGCGCGCGCAAGAGCGCTGCTGGAAGTGGCGCAAGAGGCC 1248  
DB 401 ThrAlaArgAsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGlnGly 420  
QY 1249 CACGAGATGAAGATGACGACCGAGCGCGCGCAAGCTTCCTGGGCAAGATCTGGCGCAGC 1308  
DB 421 HisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlnLysIleTrpProSer 440  
QY 1309 CACAAAGGCGCGCGCGCAACTTCTGCAAGAACCGGACGAGCGCGCGCGCGCGCGT 1368  
DB 441 TyrLysGlyArgProGlnLysAsnPheLeuGlnSerArgProGlu----- 454  
QY 1369 CCCACGCGCGCGCGCGCGAGAGCTTCGCG---TTGAGAGACACACCCCGCGCGCGC 1422  
DB 455 ProThrAlaProProGlnGlnGlnSerPheArgSerGlyValGlnThrThrThrProProGln 474  
QY 1423 AAGCAGAGAGCCCAAGAGCGCGAGCGCTTACCGGAGCGCGCGCGCGCGCGCGCTG 1482  
DB 475 LysGlnGlnProIleAspLysGlnLeuTyr-----ProLeuThrSerLeuAspSerLeu 492  
QY 1483 TTCGCGAGCGCGCGCGCTGAGCGCAG 1506  
DB 493 PheGlyAsnAspProSerSerGln 500

RESULT 12  
US-09-952-060-33  
; Sequence 33, Application US/09952060  
; Patent No. 6733993  
; GENERAL INFORMATION:  
; APPLICANT: Emlint, Emilio A.

APPLICANT: Youli, Rima  
APPLICANT: Betz, Andrew J.  
APPLICANT: Chen, Ling  
APPLICANT: Kaslow, David C.  
APPLICANT: Shiver, John W.  
APPLICANT: Toner, Timothy J.  
APPLICANT: Casimiro, Danilo R.  
TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND  
FILE REFERENCE: 20747Y  
CURRENT APPLICATION NUMBER: US/09/952,060  
CURRENT FILING DATE: 2001-09-14  
PRIORITY APPLICATION NUMBER: PCT/US01/28861  
PRIORITY FILING DATE: 2001-09-14  
PRIORITY FILING DATE: 2001-09-07  
PRIORITY FILING DATE: 2001-09-07  
PRIORITY FILING DATE: 2001-03-27  
PRIORITY FILING DATE: 2000-09-15  
PRIORITY FILING DATE: 2000-09-15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 33  
LENGTH: 493  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: cpa-gag fusion open reading frame  
US-09-952-060-33

Alignment Scores:  
Pred. No.: 2,296-119 Length: 493  
Score: 2094.00 Matches: 356  
Percent Similarity: 90.53% Conservative: 34  
Best Local Similarity: 83.37% Mismatches: 31  
Query Match: 73.73% Indels: 14  
DB: 4 Gaps: 5

US-09-475-704A-4 (1-1509) x US-09-952-060-33 (1-493)

QY 100 CTGATGTGGGCGAGCGCGAGTGGAGGCTTCGCGCTGAACCCCGCGCTGTGAGACC 159  
DB 27 ILevalTrpAlaSerArgGlnLeuGlnArgPheAlaValAsnProGlyLeuLeuGlnThr 46  
QY 160 GCCGAGGCTGCAAGCAGATCATGAGCAGCTGACGCGCGCGCGCGCGCGCGCGCAG 219  
DB 47 SerGlnGlyCysArgGlnIleLeuGlnGlnLeuGlnProSerLeuGlnThrGlySerGln 66  
QY 220 GAGCTGCGGAGCTGTGACACACCGTGGCCACCTGTACTGGCTGACACCGCGCATCGAG 279  
DB 67 GluLeuArgSerLeuLysThrValAlaThrLeuTyrCysValHisGlnLysIleAsp 86  
QY 280 GTCCGCGACACCAAGAGGCGCGTGGACAGATCGAGAGAGAGAGAGAGAGAGAGAGAG 339  
DB 87 ValLysAspThrLysGlnAlaLeuGlnLysIleGlnGlnGlnGlnGlnGlnGlnGlnGln 106  
QY 340 AAGACCCAGCAGCGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 390  
DB 107 LysAlaGlnGlnAlaAlaAlaGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrPro 126  
QY 391 ATCGTCGAGAACTTGCAGGCGCGAGATGTGTCACACGAGCCATCAGCGCGCGCGCGAGC 450  
DB 127 ILevalGlnAsnLeuGlnGlnGlnMetValHisGlnAlaIleSerProArgThrLeuAsn 146  
QY 451 GCTGAGTGAAGTATCGAGAGAGAGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGC 510  
DB 147 AlaTrpValLysValValGlnGlnLysAlaPheSerProGlnValIleProMetPheSer 166  
QY 511 GCCCTGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 570  
DB 167 AlaLeuSerGlnGlyValAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGly 186

QY	577	CAC	CAGGCGCCGATGAGATGCTGAAGAGACACCATCAACAGAGAGCGCGGATGGAC	630
Db	187	H	AGlnAlaAlaMetGlnMetLeuYbGlnThrTlleasnGlnGlnAlaAlaGlnTrrPheP	206
QY	631	CGC	CTGACCCCGGCGAGGCGCGGCGCGTGGCGCCCGGCGAGTGGCGACCCCGCGCG	690
Db	207	AG	GLeuHlaPProValHlaAlaGlyProIleAlaProGlyGlnMetArgGlnProArgGly	226
QY	691	AGG	CATGGCGGGCGCACCGACCCCTGGACAGACATGCCCTGGATGACACGAC	750
Db	227	Se	AspIleAlaAlaGlyThrThrSerThrLeuGlnGlnGlnIleGlyTrrPmetTrnaAsn	246
QY	751	CCCC	CGCGCGCGCGGCGGCGCATCTCAAGCGGTGATCATCTGGAGCGTGAACAAGATC	810
Db	247	Pro	ProIleProValGlyGlnGlyTrrYbArgTrrIleIleLeuGlyLeuAsnYbSile	266
QY	811	GTC	CGGATGTACAGCCCGGTGACATCTTGACATCGCGCGGCGCCCAAGAGCCCTTC	870
Db	267	Val	ArgMetTyrSerProThrSerThrIleLeuAspIleArgGlnGlyProYbGlnProPhe	286
QY	871	CGG	GACTAGTGGACCGGCTTTCTTCAAGACCCCTGGCGGCGGAGAGGCGACCCAGACGTG	930
Db	287	Arg	AspArgTyrValAspArgPheTyrTrrYbThrLeuArgAlaGlnGlnAlaSerGlnGlnVal	306
QY	931	AAGA	CTGGATGACCGACGACCCCTGGCGGCGAAGAGCCAAACCCGCGATGCAAGACATC	990
Db	307	lys	AsnTrrPmetThrGlnThrLeuLeuValGlnAsnAlaAsnProAspCysYbTrrIle	326
QY	991	CTG	CGCGCTCTCGGCGCGCGCGCGCACCTTGAGAGATGATGACCGCTTCCAGGCGGTG	1051
Db	327	Leu	ysAlaLeuGlyProAlaAlaThrLeuGlnGlnMetThrAlaCysGlnGlyVal	346
QY	1051	GCG	CGCGCGCGCGCGCACAGGCGCGGCTGGCGCGAGGCGATGACCG--GCCAACAGC	1101
Db	347	Gly	GlyProGlyYbIleYbAlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnSer	366
QY	1108	GTGA	CATCATATGATGTCAGAGAAGACCACTCAAGGGCGCGCGGCGCAACGTCAAGTCTTC	1168
Db	367	Ala	trIleMetLeuGlnArgGlyAsnPheAlaAsnGlnArgYbThrValIlyCysPhe	386
QY	1168	AACT	GCGGCAAGAGGGCGCACATCGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGGCTGC	1228
Db	387	Asn	CysGlyYbIlyValGlyIleHisIleAlaYbAsnCybArgAlaProArgYbYbGlyCys	406
QY	1228	TGG	AGTGGCGCAAGAGGGCGCACCATGAAAGATGCAACCGAGCGCGCAAGCGCACTTC	1288
Db	407	Trr	YbCysGlyYbIlyGlnGlyYbIleGlnMetYbAspCybAsnGlnIlyArgIleAlaAsnPhe	426
QY	1288	CTG	GCGCAAGATTTGGCGCGCGCACCAAGAGGGCGCGCGCAACTTCTCGTGAAGAAGCGCGAC	1348
Db	427	Leu	GlyYbIleTrrProSerHisIlyYbGlyArgProGlyAsnPheLeuIleAsnArgPro	446
QY	1348	GAG	CGCGCGCGCGCACCGGTGCCACCGCGCGCGCGCGCGCGAGAGCTTCGCGTTC-----	1408
Db	447	Gln	-----ProThrAlaProProGlnGlnGlnSerPheArgPheGlyGln	460
QY	1402	GAG	GAGACCAACCCCGCGCGCGCGCACAGAGAGGCCAAAGAACCGGAGCGCTTACCGGACGCC	1462
Db	461	Gln	YbTrrTrrProSerGlnIlyGlnGlnProIleAspYbGlnIleuTyr-----Pro	478
QY	1462	CTGA	CGCGCGCGCGCGCGCTGTGGCGAGCGCGCGCGCGCGAGCGAG	1506
Db	479	Leu	AlaSerLeuArgSerIleuPheGlyAsnAspProSerSerGln	493

1 NUMBER OF SEQUENCES: 40  
 2 CORRESPONDENCE ADDRESS:  
 3 ADDRESSEE: Foley & Lardner  
 4 STREET: 3000 K Street, N.W., Suite 500  
 5 CITY: Washington  
 6 STATE: D.C.  
 7 COUNTRY: USA  
 8 ZIP: 20007-5109  
 9 COMPUTER READABLE FORM:  
 10 MEDIUM TYPE: Floppy disk  
 11 COMPUTER: IBM PC compatible  
 12 OPERATING SYSTEM: PC-DOS/MS-DOS  
 13 SOFTWARE: PatentIn Release #1.0, Versid  
 14 CURRENT APPLICATION DATA:  
 15 APPLICATION NUMBER: US/08/127,499A  
 16 FILING DATE: 28-SEP-1993  
 17 ATTORNEY/AGENT INFORMATION:  
 18 NAME: BENT, Stephen A.  
 19 REGISTRATION NUMBER: 29,768  
 20 REFERENCE/DOCKET NUMBER: 51916/102/INB1  
 21 TELECOMMUNICATION INFORMATION:  
 22 TELEPHONE: (202)672-5300  
 23 TELEFAX: (202)672-5399  
 24 TELEX: 904136  
 25 INFORMATION FOR SEQ ID NO: 11:  
 26 SEQUENCE CHARACTERISTICS:  
 27 LENGTH: 478 amino acids  
 28 TYPE: amino acid  
 29 STRANDEDNESS:  
 30 TOPOLOGY: unknown  
 31 US-08-127-499A-11

US-08-127-499A-11

Pred. No.:	8.57e-115	Length:	4
Score:	2018.50	Matches:	3
Percent Similarity:	87.88%	Conservative:	4
Best Local Similarity:	78.60%	Mismatches:	1
Query Match:	71.07%	Indels:	1
DB:	1	Gaps:	5

US-09-475-704A-4 (1-1509) X US-08-127-499A-11 (1-478

[illegible]

472 GAGAGGCTTCAGCCCGAGGTGATCCCATGTTCAACCCGCTGAGGAGGCGCCACC 531  
Db 161 GIUUYAAlaPheSerProGluValIleProMetPheSerIleLeuSerIleValIleThr 180  
Qy 532 CCCGAGCCTTGAAACAGATGTTGAACACCGTGCGCGCCACGAGCCGCACTGAGATG 591  
Db 181 ProGlnAblPheAenThrMetLeuAenThrValIleGlyIleGlnIleAlaIleMetGlnMet 200  
Qy 592 CTGAAGAACCATTCAACGAGAGGCGCCGAGTGAGACCCGCTGACCCCGCTGACAGGC 651  
Db 201 LeuYsGluThrIleAenGluValIleAlaGluTProAArgValIleProValIleAla 220  
Qy 652 GGGCCCGGCGCCCGCGAGATGGCGACCCCGCGAGAGAGATGGCGCGCCACC 711  
Db 221 GlyProIleAlaProGlyGlnMetArgGluProAArgGlySerAspIleAlaGlyThr 240  
Qy 712 AGCAACCTGAGAGACAGATCGCTGATGACCAACACCCCGCTGCGCGCTGAGCGAC 771  
Db 241 SerThrLeuGlnGluGlnIleGlyTTrMetThrAenAenProIleProValIleGlu 260  
Qy 772 ATCTACAAGCGGTGATCATCTGAGGCTGAAACAAGATGTCGAGATGACAGCCCGTG 831  
Db 261 IleTyrIleArgTTrIleIleGluGlyLeuAenIleValIleArgMetTyrSerProThr 280  
Qy 832 AGCATCTGAGACATCGCGCCAGGCGCCCAAGAGCCCTTCGCGACTAGTGAACCGCTTC 891  
Db 281 SerIleLeuAspIleArgGlnIleProGlyGluProPheArgAspTyrAlaAspArgPhe 300  
Qy 892 TTCAGACCTGCGCGCGAGAGAGCCACCCGAGCTGCAAGACATCTGCGCGCTTCGCGCG 951  
Db 301 TyrIleThrLeuArgIleGluGlnIleAenGlnIleValIleAenThrMetThrGluThr 320  
Qy 952 CTGCTGTGTCAGAACGCCAACCCCGAGCTGCAAGACATCTGCGCGCTTCGCGCGCG 1011  
Db 321 LeuLeuValIleGlnAenAlaAenProAArgCysIleThrIleLeuValIleLeuIleProAla 340  
Qy 1012 GCCACCTGAGAGATGATGACCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCAAGGCC 1071  
Db 341 AlaThrLeuGlnGluMetMetThrAlaCysGlnIleValIleGlyProGlyIleValIle 360  
Qy 1072 CGCGGTGTCGCGAGGCGAGTGAAGCCAG--GCCAACAGCGTGAACATCATGATGAGAG 1128  
Db 361 ArgValIleLeuAlaGluIleAenSerGlnValIleThrAenThrAlaThrIleMetMetGlnArg 380  
Qy 1129 AGCAACTTCAAGGCGCGCGCGCAACGTCAAGTCTTCAACTGCGCGCAAGAGGCGCAC 1188  
Db 381 GlyAenPheArgAenGlnArgIleGlyMetValIleCysPheAenCysGlyIleGluGlyHis 400  
Qy 1189 ATCGCCAAAGACTGCGCGCGCGCGCGCAAGAGGCGTGTGGAAGTGGCGCAAGAGGCG 1248  
Db 401 ThrAlaArgAenCysArgAlaProArgIleGlyCysTTrIleCysGlyIleGluGly 420  
Qy 1249 CACGAGATGAGAGATGAGCGAGCGCGCAACTTCTGCGAGAGATC---TGCGCGC 1305  
Db 421 HisGlnMetIleValPheCysThrGluGlnIleAenPheLeuGlyIleCysLeuPro 440  
Qy 1306 AGCAACAGGCGCGCGCGCAACTTCTGCAAGACCGAGAGGAGCCGCGCGCGCGCAC 1365  
Db 441 ThrArgGluGlyGlnIleIlePhePheArgAlaAenGlnSerGln----- 455  
Qy 1366 GTGCCACCGCGCGCGCGCGAGAGCTTCGCTTCAAGAGAGACACCCCGCGCGCGCAAG 1425  
Db 456 -----GlnProHisIlePhePheArgAlaAenGlnSer----- 466  
Qy 1426 CAGAGCCCAAGAGACCGC 1443  
Db 467 GlnGlnProHisGlnIle 472

RESULT 14  
US-08-482-847-11  
; Sequence 11, Application US/08482847  
; Patent No. 5556757

GENERAL INFORMATION:  
APPLICANT: VAN ALSTYNE, Diane  
TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR  
TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR  
TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,847  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/127,499  
FILING DATE: 28-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 51916/104/INBI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 478 amino acids  
TYPE: amino acid  
STRANDS: 1  
TOPOLOGY: unknown  
US-08-482-847-11  
Alignment Scores:  
Pred. No.: 8.57e-115 Length: 478  
Score: 2018.50 Matches: 382  
Percent Similarity: 87.86% Conservative: 45  
Best Local Similarity: 78.60% Mismatches: 40  
Query Match: 71.07% Indels: 19  
DB: 1 Gaps: 5  
US-09-475-704a-4 (1-1509) x US-08-482-847-11 (1-478)  
Qy 1 ATGGGCGCGCGCGCGAGCATCTGCGCGCGAGAGCTGACAAAGTGAGAGATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyIleGluLeuAenArgTTrGluIleArg 20  
Qy 61 CTGCGCGCGCGCGAGAGACACTACTGCTGAAGCACTGTGTGGCCAGCCGCGAG 120  
Db 21 LeuArgProGlyGlyIleValIleGlyTyrIleLeuYsHisIleValIleThrAlaSerArgIle 40  
Qy 121 CTGAGAGGCTTGGCGCTGAACCCCGCGCTGTGAGACCGCGAGGCGCTGACAGATC 180  
Db 41 LeuGluArgPheAlaValAenProGlyLeuLeuGluTTrSerGluIleCysArgGlnIle 60  
Qy 181 ATGAAGCAGCTGACCGCGCGCTGACAGACCGAGACCGAGAGAGTGGCGAGCTGTACAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrIleSerGluGluIleLeuAenSerLeuTyrAen 80  
Qy 241 ACCGTGCGCACCTGTACTGCTGTCGACCGCGCATTCAGTCCGCGACACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValIleGlnArgIleGluIleTyrAspThrIleGluIle 100  
Qy 301 CTGACAAGATGACGAGAGAGACAGAACAAAGTCCACGAGAAAGACCCAGCGCAAGAG 360

[illegible]

```

Db      456-----GlnProHisHisPhePheArgAlaAspGlnSer-----466
QY      1426 CAGGAGCCCAAGACCCG 1443
      |||::|||
      467 GlnGlnProHisGlnIys 472

RESULT 15
US-08-392-794A-2
; Sequence 2, Application US/08392794A
; Patent No. 6025141
; GENERAL INFORMATION:
; APPLICANT: HO, Yu-Wen
; TITLE OF INVENTION: IMMUNOFLOURESCENCE ASSAY FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES USING RECOMBINANT ANTIGENS IN INSOLUBLE FORM
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/392,794A
; FILING DATE: 09-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/164,789
; FILING DATE: 10-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: NEIMARK, SHERIDAN
; REGISTRATION NUMBER: 20,520
; REFERENCE/DOCKET NUMBER: HU=4A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-392-794A-2

Alignment Scores:
Pred. No.: 2,41e-114 Length: 437
Score: 2011.00 Matches: 374
Percent Similarity: 93.14% Conservative: 33
Best Local Similarity: 85.58% Mismatches: 26
Query Match: 70.81% Indels: 4
DB: 3 Gaps: 2

US-09-475-704A-4 (1-1509) x US-08-392-794A-2 (1-437)
QY      1 ATGGGCGCGCCGCGCAGCATCTGCGCGCGGAGGAGTGAAGTGAAGATGATCCG 60
      |||||::|||
Db      1 MetGlyAlaArgAlaSerValIleuSerIyGlyIuLeuAbaPgrIpoluylSer 20
QY      61 CTGGCGCCCGCGCGGAGAGACCTACATGCTGAAGCAGCTGTGTGGGCGAGCCGCGAG 120
      |||||::|||
Db      21 LeuArgProGlyIyLysIyLysIyLysIyLysIyLysIyLysIyLysIyLysIy 40
QY      121 CTGAGAGGAGCTTGGCCCTTGAACCCCGGCGCTGTGGAGACCGCGAGGGCTGCGAAGAGATC 180
      |||||::|||
Db      41 LeuGlnArgPheAlaValIleuProGlyLeuLeuGlnIuTrSerGlnGlyCyArgGlnIle 60
      |||||::|||
QY      181 ATGAAGAGAGCTGACCGCCGCTTGACAGCCGACCGAGGAGCTGGCAGCCTGTACAAC 240

```



**THIS PAGE BLANK (USE TO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 14:56:14 ; Search time 206.806 Seconds

(without alignments)  
5644.151 Million cell updates/sec

Title: US-09-475-704A-4

Perfect score: 2840

Sequence: 1 atggcgccgcgcgcgcagcat.....gcgcgcctcgtgacagctaa 1509

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
	Ygapop 10.0 , Ygapext 0.5
	Fgapop 6.0 , Fgapext 7.0
	Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-O=/cn2.1/USPTO.spool.p/US09475704/runat.10032005.140220.14808/app.query.fasta\_1.3342  
-DB=A.Geneseq.16Dec04 -OPMT=fastaan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biolum62 -TRANS=humand0.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704 @CGN 1.1 472 @runat.10032005.140220.14808 -NCPU=6 -ICPU=3  
-NO MMAP -LARGECQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -NARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.Geneseq.16Dec04:\*

1: geneseqp1980s:\*\n2: geneseqp1990s:\*\n3: geneseqp2000s:\*\n4: geneseqp2001s:\*\n5: geneseqp2002s:\*\n6: geneseqp2003as:\*\n7: geneseqp2003bs:\*\n8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2661	93.7	502	3	AA96944
2	2464	86.8	492	6	AA37600
3	2463	86.7	492	8	ADP02848
4	2463	86.7	513	8	ADP02847
5	2457	86.5	492	5	AA318303
6	2457	86.5	635	5	AA448948
7	2440	85.9	492	7	ADC72879
8	2430	85.6	492	7	ADC72880
9	2421.5	85.3	491	7	ADC72910
10	2415.5	85.1	1457	8	ADN36406

11	2413.5	85.0	491	7	ADC72913
12	2410.5	84.9	514	8	ADN36396
13	2404	84.6	498	3	AA869276
14	2402	84.6	494	3	AA869275
15	2387	84.0	508	7	ADC72875
16	2385.5	84.0	491	7	ADC72874
17	2381.5	83.9	491	3	AA869278
18	2375.5	83.6	497	7	ADC72877
19	2353.5	82.9	487	7	ADC72876
20	2348	82.7	496	7	ADC72878
21	2343	82.5	492	3	AA96943
22	2328.5	82.0	3025	4	AA866196
23	2273	80.0	494	3	AA869271
24	2270	79.9	500	1	AA80884
25	2258.5	79.5	1229	6	AA37597
26	2258	79.5	500	7	ADC72788
27	2253	79.3	937	2	AA366822
28	2252.5	79.3	497	2	AA808040
29	2251	79.3	442	6	AA37598
30	2247.5	79.1	494	7	ADC72896
31	2247	79.1	502	2	AAW53108
32	2247	79.1	502	3	AA77294
33	2247	79.1	502	6	ABU63178
34	2247	79.1	506	1	AA660176
35	2247	79.1	506	2	AA829704
36	2247	79.1	506	5	AA35783
37	2246	79.1	500	2	AAW63737
38	2246	79.1	500	3	AA810046
39	2246	79.1	500	3	AA70599
40	2246	79.1	500	3	AA770598
41	2246	79.1	500	6	ABP96542
42	2246	79.1	500	6	ABR82426
43	2246	79.1	500	7	ABR92015
44	2246	79.1	500	8	AD128568
45	2246	79.1	512	8	AD128569

## ALIGNMENTS

RESULT 1  
AA96944  
ID AA96944 standard; protein; 502 AA.  
XX  
AC AA96944;  
XX  
DT 31-OCT-2000 (first entry)  
XX  
DE HIV synthetic Gag containing polypeptide.  
XX  
KW Gag; expression cassette; antigenic; type C; HIV; Env; synthetic;  
XX DNA immunization; packaging cell line; antigen presentation.  
XX  
OS Human immunodeficiency virus; type C strain AF110967.  
OS Synthetic.  
OS  
PN WO200039304-A2.  
XX  
PD 06-JUL-2000.  
XX  
PF 30-DEC-1999; 99WO-US031273.  
XX  
PR 31-DEC-1998; 98US-0114495P.  
XX  
PR 01-SEP-1999; 99US-0152195P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Barnett S, Zur Megede J;  
XX  
DR WPI; 2000-452401/39.  
XX  
PT Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env  
PT polypeptide and the polypeptide useful for immunizing a mammal especially  
PT human against HIV.





RESULT 3  
ADP02848  
ID ADP02848 standard; protein; 492 AA.  
AC ADP02848;  
XX  
XX  
DT 18-NOV-2004 (first entry)  
DE Wild type HIV-1 gag protein from strain Du499.  
XX  
XX anti-HIV; vaccine; HIV; Gag; prophylaxis; immunogenic response.  
KM  
OS Human immunodeficiency virus 1.  
XX  
XX WO2004050691-A2.  
XX  
XX 17-JUN-2004.  
XX  
XX 04-DEC-2003; 2003WO-1B005634.  
XX  
XX 04-DEC-2002; 2002ZA-00009830.  
XX  
XX (UYCA-) UNIV CAPE TOWN.  
XX (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.  
XX  
XX Jaffray A, Williamson A, Rybicki EP;  
PI WPI; 2004-450716/42.  
XX  
XX N-PSDB; ADP02846.  
DR  
XX  
XX New vector including a nucleotide sequence encoding the Gag polypeptide,  
PT useful as a vaccine for the treatment or prophylaxis of HIV infection in  
a mammal.  
XX  
XX Disclosure; SEQ ID NO 4; 37bp; English.  
XX  
XX The invention relates to a vector including a nucleotide sequence  
CC encoding an HIV Gag polypeptide, where the nucleotide sequence encoding  
CC the Gag polypeptide comprises a sequence having at least 90% sequence  
CC identity to 1549 (S1) or 1479 (S2) bp defined in the specification. The  
CC vector is useful as a vaccine for the treatment or prophylaxis of HIV  
CC infection in a mammal, where the vaccine induces an immunogenic response  
CC to the virus-like particles in a suitable susceptible host. This sequence  
CC corresponds to an HIV-1 gag protein sequence used in the method of the  
invention.  
XX  
XX  
SQ Sequence 492 AA;  
XX  
XX  
Alignment Scores:  
Pred. No.: 3,33e-132 Length: 492  
Score: 2463.00 Matches: 466  
Percent Similarity: 95.42% Conservative: 13  
Best Local Similarity: 92.83% Mismatch: 13  
Query Match: 86.73% Indels: 10  
DB: 8 Gaps: 2  
US-09-475-704a-4 (1-1509) x ADP02848 (1-492)  
QY 1 ATGGGCGCCCGCGCGCATCTGCGCGCGGAGAGACTGAGCAAGTGGAGAAATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCCCGCGCGGAGAGAGACTGAGTGGAGCACTGAGTGGCGCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysHisTrpMetLeuLysHisIleValTrpAlaSerArgLys 40  
QY 121 CTGGAGGGCTTGGCGCGCGCGCGCTGCTGAGAGACCGGAGGGCTGCAAGCATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60  
QY 181 ATGAAGACAGCTGAGCGCGCGCGCTGCAAGCGGAGCGGAGAGTGGCGAGCTGTACAC 240  
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLysSerLeuLysTrpAsn 80

QY 241 ACGGTGGCAACCTGTACTGTGTCAGCGCGGCATCGAGGTCCGCGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuLysCysValHisGluLysIleGluValAlaArgPheThrLysGlnAla 100  
QY 301 CTGGAACAAGATCGAGAGAGACAGAACAGTCCCGACAGAGACCCAGCGCCAGAG 360  
DB 101 LeuAspLysIleGluGluGlnGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCCGACGCGAAGTGTAGCGAGACTACCCCATCTGTCGAAACCTGCAAGGCGCAATG 420  
DB 121 AlaAspGlyLysValSerGlnAsnLysProIleValGlnAsnLeuGlnGlyGlnMetVal 140  
QY 421 CACGAGGCGATCAACCGCGCGCGCTGAGAGCGCTGGAGAGTGTATGAGAGAGGCC 480  
DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluValAla 160  
QY 481 TTCAGCCCGAGTGTATCCCATGTTCACCGCGCTGAGCGAGGCGCCACCCCGAGAC 540  
DB 161 PheSerProGluValIlePheMetPheThrAlaLeuSerGluGluAlaThrProGlnAsp 180  
QY 541 CTGAACACAGATGTGAACACCGTGGCGCGGCCACAGCGCGCGCATATGCTTAAGAC 600  
DB 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAAGAGAGAGCGCGCGAGTGGGACCGCGCTGACCCCGTGCAGGCGCGCGCTG 660  
DB 201 ThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
QY 661 GCCCGCGCGAGATGCGGAGCCCGCGCGCGACGCAATCGCGCGCGCCACGACCTGT 720  
DB 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGGAGCAGATCGCTGTGATATACAGCAACCCCGCGCGCGGAGGAGCATATCAAG 780  
DB 241 GlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleLys 260  
QY 781 CGGTGATCATCTGTGGCGCTGAACAAGTCTGCGAGTGTACACCCCGTACAGATCTGT 840  
DB 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetCysSerProValSerIleLeu 280  
QY 841 GACATCGCGAGGCGCGCGAGAGCGCTTCCGCGCATCTGAGAACCGCTTTCAAGACC 900  
DB 281 AspIleArgGlnGlyProLysGluProPheArgAspTrpValAspArgPhePheLysThr 300  
QY 901 CTGGCGCGCGAGCGCGCCACGAGCACTGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 960  
DB 301 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuLeuVal 320  
QY 961 CAGAAAGCGCAACCGCGAGTGAAGACATCTGCGCGCTCTCGGCGCGCGCGCGCACCTG 1020  
DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340  
QY 1021 GAGGAGATGATGACCGCTTGCAGAGGCGCTGGCGCGCGCGCGCGCACAGGCGCGTGT 1080  
DB 341 GlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlnHisLysAlaArgValLeu 360  
QY 1081 GCCGAGGGGATGAGCGAGCGCAAGCGGTGAACATCATGATGCAAGAGCAACTTCAAG 1140  
DB 361 AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetCysGlnLysSerAsnPheLys 380  
QY 1141 GCGCGCGCGCGCAACGTCAATGTCTTCAATCTGCGCGAGAGGCGCGCATGCGCGCAAGAC 1200  
DB 381 GlyProAlaGlyIleValLysCysPheAsnLysCysGlyLysGluGlnHisIleAlaArgAsn 400  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGTGGCGCAAGAGGCGCGCAATGTAAG 1260  
DB 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnHisGlnMetLys 420  
QY 1261 GACTGCAACGAGCGCGCGCGCAACTTCTGGGCAAGATCTGAGCGCGCAAGAGGCGCGC 1320  
DB 421 AspCysThrGluArgGlnAlaAsnPheLeuGluLysIleTrpProSerHisLysGlyArg 440

```

QY      1331  CCGCGCACTTCTCTGAGAAACCGCAGCGCCGCCGCCACCGTCCACCGCCCC 1380
Db      441  ProGlyAmpPheLeuGlnAsnArgProGlu-----ProThrLapPro 454
QY      1391  CCGCGGAGAGCTTCGGCTTGGAGAGACCAACCGCGCCGCCCAAGAGAC 1440
Db      455  ProLlaGluSerPheArgPheGluGluTrnTrnProLapProLysGlnGluProLleGlu 474
QY      1441  CGCGAGCCCTTACCGCGAGCCCTGTACCGCCCTGTGCGAGCTGTTCGAGCGGCCCTGT 1500
Db      475  -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490
QY      1501  AGCCAG 1506
Db      491  SerGln 492

RESULT 4
ADP02847
ID      ADP02847 standard; protein; 513 AA.
XX      ADP02847;
XX      18-NOV-2004 (first entry)
XX      Wild type HIV-1 gag protein from strain Du499.
XX      anti-HIV; vaccine; HIV; Gag; prophylaxis; immunogenic response.
XX      Human immunodeficiency virus 1.
XX      WO2004050691-A2.
XX      17-JUN-2004.
XX      04-DEC-2003; 2003WO-1B005634.
XX      04-DEC-2002; 2002ZA-00009830.
XX      (UYCA-) UNIV CAPE TOWN.
XX      (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.
XX      Jaffray A, Williamson A, Rybicki EP,
XX      WPI; 2004-450716/42.
XX      N-PSDB; ADP02845.
XX      New vector including a nucleotide sequence encoding the Gag polypeptide,
XX      useful as a vaccine for the treatment or prophylaxis of HIV infection in
XX      a mammal.
XX      Disclosure; SEQ ID NO 3; 37bp; English.
XX      The invention relates to a vector including a nucleotide sequence
XX      encoding an HIV gag polypeptide, where the nucleotide sequence encoding
XX      the gag polypeptide comprises a sequence having at least 90% sequence
XX      identity to 1549 (S1) or 1479 (S2) bp defined in the specification. The
XX      vector is useful as a vaccine for the treatment or prophylaxis of HIV
XX      infection in a mammal, where the vaccine induces an immunogenic response
XX      to the virus-like particles in a suitable susceptible host. This sequence
XX      corresponds to an HIV-1 gag protein sequence used in the method of the
XX      invention.
XX      Sequence 513 AA;

Alignment Scores:
Pred. No.:      3,346-132      Length:      513
Score:          2463.00      Matches:      466
Percent Similarity: 95.42%      Conservative: 13
Best Local Similarity: 92.83%      Mismatches:  13
Query Match:    86.73%      Indels:      10
DB:            Gaps:          2

US-09-475-70A-4(1-1509) x ADP02847 (1-513)

```

QY	1	ATGGAGCGCCGCGCCAGCATCTTCGCGGCGAGAACTGGACAACTGGAGAGAAATCCG	60
DB	3	MetGlyAlaArgAlaSerIleLeuArgGlyGluIysLeuAspLysTrpGluLysIleArg	22
QY	61	CTGGGCGCCCGCGGCAAGAGACATTAATGCTGAAGACCTCGGTGGCGCCAGCGGAG	120
DB	23	LeuArgProGlyGlyLysLysHisArgTyrMetLeuLysHisIleValTrpAlaSerArgGlu	42
QY	121	CTGAGAGGAGCTTCGCGCTGAACCCCGGCGCTGAGACCCGCGAGGCGTGCAGACAGATC	180
DB	43	LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysArgGluIle	62
QY	181	ATGAAGACGCTGACGCCCGCCCTTCGACACCGGACCGAGAGGCTGGCCGACGCTTACAC	240
DB	63	MetLysGluLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuLysSerLeuTyrAsn	82
QY	241	ACCGTGGCCACCTGTACTGCGTGCACGCGCGGACTGAGAGTCCGCGACCAACAGAGGCG	300
DB	83	ThrValAlaThrLeuTyrCysValHisGluLysIleGluValAlaArgAspThrLysGluAla	102
QY	301	CTGGAACAAGATCGAGAGAGAGACAGAACAGTCCACAGAAAGACCCAGAGGCCAAGAG	360
DB	103	LeuAspLysIleGluGluGluGlnAsnLysCysGlnGluLysThrGlnGlnAlaLysAla	122
QY	361	GCCGACGGGCAGAGGTAGCGAGAACTACCCCATCGTGAAACCGAGAGGCGCAATGCTG	420
DB	123	AlaAspGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyMetVal	142
QY	421	CACCAAGGCATCAGCCCGCGCACCTGTAGCGCTGGTGAAGTATCGAGAGAGGCC	480
DB	143	HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla	162
QY	481	TTTCAGCCCCGAGGTGATTCCTCCATGTTCAACGCGCTTGAGCCAGAGGCGCACCCCGCAGAC	540
DB	163	PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGluAsp	182
QY	541	CTGAACAACGATGTTGAACAACGCTGGGCGGCGCACAGGCGCGCATGCAGATCTGAAGAC	600
DB	183	LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp	202
QY	601	ACCATCAACGAGAGGCGCGCGAGTGGAGACCGCTCGCAACCCCGTGACAGCGCCCGCTG	660
DB	203	ThrIleAsnGluGlnAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProIle	222
QY	661	GGCCCGCGCGCAGATGCGCGACCCCGGCGGAGCAATCGCGCGGCGCACGAGACCTGT	720
DB	223	AlaProGlyGlnMetArgLysProArgLysSerAspIleAlaGlyThrThrSerThrLeu	242
QY	721	CAGGAGCAGATCGCGCTGATGACACGAGACCCCGCGTGGCGCCGATCTACAG	780
DB	243	GlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTyrLys	262
QY	781	CGGTGATCATCTCGGCGCTGACAACAATGTTGGGATGTTACAGCCCGCTGAGATCTGT	840
DB	263	ArgTrpIleIleLeuGlyLeuAsnLysAlaValaArgMetLysSerProValSerIleLeu	282
QY	841	GACATCGCGCAGGCGCCCAAGAGACCTTCGCGACATCAGTGGACCGCTTCTTAAGAC	900
DB	283	AspIleArgGlnGlyProLysGluProPheArgAspTyrValaAspArgPhePheLysThr	302
QY	901	CTGCGGCGCGAGCAGCGCCACCCAGAGCGTGAAGAACTGATGACCGAGACCTGTGCTG	960
DB	303	LeuArgAlaGluGlnAlaTrpGlnGluValLysAsnTrpMetThrAspThrLeuVal	322
QY	961	CAGAAAGCGCAACCCCGCATGGAAGACATCTGCGGCGCTCGCGCGCCGCGGACACCTG	1020
DB	323	GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu	342
QY	1021	GAGGAGATGATGATCGGCTGCGCAGAGGCTGGGCGGCGCCGCGCAAGCGCCGCGTCTG	1080
DB	343	GluGluMetMetThrAlaCysGlnGlyValGlyValProGlyHisLysAlaArgValLeu	362

```
QY 1081 GCCGAGCGCATGAGCCAGCCAGCGTGAACATCATGATGAGAGAGCAATTCAC 1140
    |||||
Db 363 AlGlnuAlaMetSerGlnThrAsnSerGlyAsnIleMetGlnAngSerAsnPhenLeys 382
    |||||
QY 1141 GGGCCCCGGCGCAACGTCAAGTGTTCACCTGGCGGCAAGAGGCGCATCGCCAGAAC 1200
    |||||
Db 383 GlyProkArgArgIleValIysCysPheAsnCysGlyIysGlnIleAlaArgAsn 402
    |||||
QY 1201 TGCCCGCCCCCGGCAAGAGGCGTGTGATGAGTGGCGGCAAGAGGCGCCACGATGAG 1260
    |||||
Db 403 CysArgAlaProkArgIysIysGlyCysTrpIysCysGlyIysGlnIleAlaArgAsn 422
    |||||
QY 1261 GACTGCAACGAGCGCGCAACCTTCCTGGGCAAGATCGGCGCAAGAGGCGCG 1320
    |||||
Db 423 AspCysThrGlnArgIleAlaAsnPhenLeuGlyIleTrpProSerHisIysGlyArg 442
    |||||
QY 1321 CCGCGCACTTCCTGCAAGACCGCAAGAGCGCGCGCGCCCGCCACGTCGCCACCGCC 1380
    |||||
Db 443 ProGlyAsnPhenLeuGlnAsnArgProGlu-----ProThrAlaPro 456
    |||||
QY 1381 CCGCGCGAGAGCTTCGCTTCGAGAGACACACCCCGCCCGCCAGCAAGAGCGCCAGAC 1440
    |||||
Db 457 ProAlaGlnSerPheArgPheGlnGlnThrProAlaProIysGlnGluProIleGlu 476
    |||||
QY 1441 CGCGAGCGCTACCGCGAGCGCGCTGACCGCGCTGCGAGCGCTTCGCGAGCGCGCTG 1500
    |||||
Db 477 -----ArgGluProLeuThrSerLeuIysSerLeuPheIysSerAspProLeu 492
    |||||
QY 1501 AGCCAG 1506
    |||||
Db 493 SerGln 494

RESULT 5
AAE18303
ID AAE18303 standard; protein; 492 AA.
XX
AC AAE18303;
XX
DT 29-AUG-2003 (revised)
DT 07-MAY-2002 (first entry)
XX
DE Human immunodeficiency virus type 1 (HIV-1) Clade C gag protein.
XX
KW Human immunodeficiency virus; vaccine; HIV infection; immune response;
KW alpha-virus replicon; therapy; gag gene.
XX
OS Human immunodeficiency virus 1.
XX
PN WO200203917-A2.
XX
PD 17-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-US021701.
XX
PR 07-JUL-2000; 2000US-0216995P.
XX
PA (ALPH-) ALPHAVAX INC.
PA (VINC-) UNIV NORTH CAROLINA.
XX
PI Olmsted R, Keith P, Dryga S, Caley I, Maughan M, Johnston R;
PI Davis N, Swannstrom R;
XX
DR WPI, 2002-171664/22.
DR N-PSDB; AAD29133.
XX
PT Composition useful for treating or preventing HIV infections, comprises
PT two or more isolated nucleic acids encoding env, gag or pol gene product
PT of HIV or immunogenic fragment of the gene products.
XX
PS Example 5; Page 165-166; 201pp; English.
XX
CC The invention relates to a composition comprising isolated nucleic acids
CC encoding env, gag or pol gene product of human immunodeficiency virus or
```

```
CC immunogenic fragment of the gene products. The gag gene product is
CC modified to inhibit formation of virus-like particles containing gag gene
CC product and their release from cells, and the pol gene product is
CC modified to inhibit reverse transcriptase activity. The invention also
CC relates to a method for producing an alpha-virus replicon particle used
CC in vaccines. The composition is useful for inducing an immune response to
CC human immunodeficiency virus (HIV) or for treating or preventing HIV
CC infection in a subject. The alpha-virus replicon particle is useful in a
CC vaccine. The composition is useful for administering a protein or peptide
CC to a subject. A composition comprising hepatitis affinity-purified alpha-
CC virus replicon particle is useful as a clinical trial material and as a
CC commercial product. The present sequence is HIV-1 Clade C gag protein.
CC (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 492 AA;

Alignment Scores:
Pred. No.: 7,31e-132 Length: 492
Score: 2457.00 Matches: 465
Percent Similarity: 95.22% Conservative: 13
Best Local Similarity: 92.63% Mismatches: 14
Query Match: 86.51% Indels: 10
DB: 5 Gaps: 2

US-09-475-704A-4 (1-1509) x AAE18303 (1-492)
QY 1 ATGGGCGCGCGCGCGCGCATCTTCGCGCGGCAAGAGTGAACAGTGGAGAGATCCGC 60
    |||||
Db 1 MetAlaAlaArgAlaSerIleLeuArgGlyIlyIysLeuAspIysTrpGlnIlyIleArg 20
    |||||
QY 61 CTGGCGCGCGCGCGCGCAAGACATACATGCTGAAGACCGTGGTGGCGCGCGCGAG 120
    |||||
Db 21 LeuArgProGlyGlyIysIysHisIleTrpMetLeuIysHisIleValIleTrpAlaSerArgGlu 40
    |||||
QY 121 CTGAGAGGCTTCGCTTCGCAACCCCGCGCTGTGAGACCGCGCGAGGCGTGCACAGATC 180
    |||||
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnIlyIysGlnIle 60
    |||||
QY 181 ATGAAGAGCGTGCAGCGCGCGCTTCGAGACCGCGCACCGAGAGCTGGCGAGCTGTACAC 240
    |||||
Db 61 MetIysGlnLeuGlnProAlaLeuGlnThrIlyThrGlnIleIysSerLeuTrpAsn 80
    |||||
QY 241 ACCGTGGCACCCGTACTGCGTGCACAGCGCGCATCGAGGTCGCGACACCAAGAGGCC 300
    |||||
Db 81 ThrValAlaThrLeuTrpCysValHisGlnIlyIleGlnValAlaGAspThrIysGlnAla 100
    |||||
QY 301 CTGACAGAGATCGAGAGAGAGACAGACAGTCCCGAGCAAGAGACCGAGCGCCAGAG 360
    |||||
Db 101 LeuAspIysIleGlnIlyGlnIlyGlnIlyAsnIysCysGlnIlyIysThrGlnIlyAlaIysAla 120
    |||||
QY 361 GCCGAGCGCAAGGAGCGCAAGTACATCCCATCGTGAAGACCTGAGCGCGCCAGATGCTG 420
    |||||
Db 121 AlaAspIlyIysValIserGlnAsnTrpIleValGlnAsnLeuGlnIlyGlnMetVal 140
    |||||
QY 421 CACGAGCGCATCAGCGCGCGCGCATCGTGAAGCGCTGGTGAAGTATCGAGAGAGGCC 480
    |||||
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValIysValIleGlnIlyIysAla 160
    |||||
QY 481 TTCAGCGCGCGAGGTGATCCCATGTTTACCGCGCTGAGCGAGGCGCCACCGCCAGGAC 540
    |||||
Db 161 PheSerProGlnValIleProMetPheThrAlaLeuSerGlnIlyAlaThrProGlnAsp 180
    |||||
QY 541 CTGAACAGAGTGTGAACACCGTGGGCGGCGACAGGCGCGCATGCGAGTGTGAAGAGAC 600
    |||||
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnIlyAlaIleMetGlnMetLeuIysAsp 200
    |||||
QY 601 ACCATCAAGAGAGAGCGCGCGAGTGGAGCGCGTGCACCGCGTGCAGGCGCGCGCGG 660
    |||||
Db 201 ThrIleAsnIlyGlnIlyAlaIlyGlnIlyIysIysIysIysIysIysIysIysIysIys 220
    |||||
QY 661 GCCCGCGCGCAAGTGGCGCGACCGCGCGGAGAGAGCATCGCGCGCGCACAGCACCTG 720
    |||||
Db 221 AlaProGlyIlyIlyMetArgIlyProArgIlySerAspIleAlaGlyIlyThrThrSerThrLeu 240
```

```

QY 721 CAGGACGATGCGCTGATGACGAGACCCCGCTGCGGCGGACATCTACAG 780
    |||
Db 241 GlnGlnGlnLeuIleAlaTrpMetThrSerAsnProProIleProAlaGlyAspIleTyrLys 260
    |||
QY 781 CCGTGATCATCTGGGCGCTGAACAAGATGTCGAGATGACGCGCGGACATCTCG 840
    |||
Db 261 ArgTrpIleIleLeuLeuIleLeuAsnLysIleValArgMetCysSerProValSerIleLeu 280
    |||
QY 841 GACATCCGCGGCGGCGCGCAAGAGCGCTTCCGACTACGTGAGCCGCTTCTTCAAGCC 900
    |||
Db 281 AspIleArgGlnGlnIleProLysGluProPheArgAspTyrValAspArgPhePheLysTrp 300
    |||
QY 901 CTGCGCGCGGCGGCGCGCAAGAGCGCTTCCGACTACGTGAGCCGCTTCTTCAAGCC 960
    |||
Db 301 LeuArgAlaGlnGlnIleAlaThrGlnGlnValLysAsnTrpMetThrAspThrLeuLeuVal 320
    |||
QY 961 CAGAACGCGCAACCCCGACTGCAAGACCATCTCGCGCTTCCGCGCGCGCGCGCGCGCG 1020
    |||
Db 321 GlnAsnAlaAsnProAspCysLysTrpIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340
    |||
QY 1021 GAGGAGATATGACCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1080
    |||
Db 341 GlnGlnMetMetThrLysAlaCysGlnGlnIleValGlyProGlyLysIleValArgValLeu 360
    |||
QY 1081 GCGGAGCGGATGAGCGCGCGCAACGCTGGAATCATGATGATGCAAGGCGCACTTCAAG 1140
    |||
Db 361 AlaGlnAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 380
    |||
QY 1141 GCGCGCGCGCGCGCAACGCTGGAATCATGCTGCGCGCGCAAGGCGCGCACTTCCGCAAG 1200
    |||
Db 381 GlyProArgArgIleValLysCysPheAsnGlyLysGlnGlyIleIleAlaArgAsn 400
    |||
QY 1201 TGGCGCGCGCGCGCGCAAGAGGCTGCTGGAATGCTGCGCGCGCAAGGCGCGCACTGAG 1260
    |||
Db 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyIleGlnMetLys 420
    |||
QY 1261 GACTGACGCGGCGCGCGCGCACTTCTGGGCAATGCTGCGCGCGCGCGCGCGCGCGCG 1320
    |||
Db 421 AspCysTrpGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
    |||
QY 1321 CCGCGCACTTCTGCGCAAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
    |||
Db 441 ProGlyAsnPheLeuGlnIleAsnArgProGlu-----ProThrAlaPro 454
    |||
QY 1381 CCGCGCGGAGGCTTCCGCTTCAAGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
    |||
Db 455 ProAlaGlnSerPheArgPheGlnGlnIleTrpProAlaProLysGlnIleProIleGln 474
    |||
QY 1441 CCGGAGCCCTACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
    |||
Db 475 -----ArgGlnProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490
    |||
QY 1501 AGCCAG 1506
    |||
Db 491 SerGln 492
    |||

```

```

FT Misc-difference 523
FT /note= "encoded by TAA"
FT Misc-difference 526
FT /note= "encoded by TAG"
FT Misc-difference 557
FT /note= "encoded by TAA"
FT Misc-difference 612
FT /note= "encoded by TGA"
XX
PN WO200204494-A2.
XX
PD 17-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-1B001208.
XX
PR 07-JUL-2000; 2000US-0216995P.
XX
PR 10-JUL-2000; 2000ZA-00003437.
XX
PR 15-SEP-2000; 2000ZA-00004924.
XX
PA (MED- ) MEDICAL RES COUNCIL.
PA (UYCA-) UNIT CAPE TOWN.
PA (UYNC-) UNIT NORTH CAROLINA.
XX
PI Williamson C, Swansstrom RI, Morris L, Karim SA, Johnston RE;
XX
DR WPI: 2002-171700/22.
XX
DR N-PSDB; AAL41591.
XX
PT Selecting HIV-1 subtype C isolates, which are useful in developing
PT vaccines against HIV infection, comprises isolating viruses with high
PT sequence identity to a consensus sequence whose phenotype is associated
PT with the HIV subtype.
XX
PS Claim 17; Page 64-66; 69pp; English.
XX
CC The present invention relates to a process for the selection of human
CC immunodeficiency virus (HIV) subtype isolates for use in the development
CC of a prophylactic and/or therapeutic pharmaceutical composition. The
CC process involves selecting isolated virus or viruses with a high sequence
CC identity to a consensus sequence and a phenotype which is associated with
CC transmission for the particular HIV subtype. The composition can be used
CC in the production of vaccines against HIV. The present sequence is the
CC HIV-1 subtype C isolate Du422 gag protein. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
SQ Sequence 635 AA;
XX
Alignment Scores:
Pred. No.: 7 45e-132 Length: 635
Score: 2457.00 Matches: 465
Percent Similarity: 95.22% Conservative: 13
Best Local Similarity: 92.63% Mismatches: 14
Query Match: 86.51% Indels: 10
DB: 5 Gaps: 2

```

```

US-09-475-704A-4 (1-1509) x AAM48948 (1-635)
QY 1 ATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
    |||
Db 31 MetAlaIleArgAlaSerIleLeuArgGlyLysLeuAspLysTrpGlnLysIleArg 50
    |||
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
    |||
Db 51 LeuArgTrpGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 70
    |||
QY 121 CTGAGGCGCTTCCGCTGAAACCCCGCGCTGAGAGCGCGCGCGCGCGCGCGCGCGCG 180
    |||
Db 71 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnIleTrpAlaSerArgGln 90
    |||
QY 181 ATGAAGCAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
    |||
Db 91 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLysSerLeuLysTrpAsn 110
    |||

```

Key Location/Qualifiers

```

QY 241 ACCGTGGCCACCTGTACTGTCTGACCGCGCATCGAGTCCGCGCAACCAAGAGGCC 300
    |||
Db 111 ThValAlaThrLeuYrCyValHsGluLysIleGluValAlaArgAspThrLysGluAla 130
QY 301 CTGGAACAAGATCGAGAGAGAGAGCAAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 360
    |||
Db 131 LeuAspLysIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 150
QY 361 GCCGACCGCAAGGTGAGCCAGAACTACCCCATGTGAGAACTGCAAGGCGCATGTG 420
    |||
Db 151 AlaAspGlySerValSerGlnAsnYrProIleValGlnAsnLeuGlnGlnMetVal 170
QY 421 CACGAGGCGCATGAGCCCGCGACCGCTGAAAGCGGTGAGTGAAGTGAAGAGAGAGCC 480
    |||
Db 171 HisGlnAlaIleSerProAspGlnThrLeuAsnAlaIrrValIleGluGluLysAla 190
QY 481 TTGAGCCCGGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCCACCCCGAGAC 540
    |||
Db 191 PheSerProGluValIleProMetPheThrAlaLeuSerGluGluValAlaThrProGlnAsp 210
QY 541 CTGGAACAGCATTTGAAACACCGTGGCGCGCGCCAGAGCGCGCATGCAATGCTGAAGAC 600
    |||
Db 211 LeuAsnThrMetLeuAsnThrValGlyGlyHsGlnAlaAlaMetGlnMetLeuLysAsp 230
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCGGCTGACCCCGTGAAGGCGCGCGCGG 660
    |||
Db 231 ThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyProIle 250
QY 661 GCCCGCGCGCGAGTGGCGAGACCCCGCGCGAGAGACATGCGCGCGCGCCAGACCCCTG 720
    |||
Db 251 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 270
QY 721 CAGGACAGATGGCTGATGACCAAGACACCCCGCTGCGCGCGCGCATCTTCAAG 780
    |||
Db 271 GlnGluGlnHisLeuAlaTrpMetThrSerAspProIleProValGlyAspIleYrLys 290
QY 781 CCGTGGATCATCTGGGCGCTGAACAAGATGTCGCGATGATGACCGCGCGCGATCTG 840
    |||
Db 291 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIleLeu 310
QY 841 GACATCCGCGAGCGCGCGCGCGAGAGCCCTTCCGCACTACGTGAGCGGCTTCTTCAAGAC 900
    |||
Db 311 AspIleArgGlnGlyProLysGluProPheArgAspYrValAlaAspArgPheLysThr 330
QY 901 CTGCGCGCGAGAGCGCGCGAGAGCGCGTGAAGAACTGATGACCGAGACCGCTGCTG 960
    |||
Db 331 LeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThrAspThrLeuVal 350
QY 961 CAGAACCGCAACCCCGCATGCAAGACATCTGCGGCTCTGCGCGCGCGCGCGCGCGCTG 1020
    |||
Db 351 GlnAsnAlaAsnProAspYrLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 370
QY 1021 GAGGAGATGATGACCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1080
    |||
Db 371 GlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu 390
QY 1081 GCCGAGCGCATGAGCGCGAGCGCGCATGATCATGATGATGATGATGATGATGATGATGAT 1140
    |||
Db 391 AlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnLys 410
QY 1141 GCGCGCGCGCGCGCAAGTCAAGTCTTCACTGCGCGCAAGAGAGAGCGCGCGCGCGAGAC 1200
    |||
Db 411 GlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 430
QY 1201 TGGCGCGCGCGCGCGAGAGAGCGCGTGGAGATGCGCGCGAGAGAGCGCGCGCGCGAGAC 1260
    |||
Db 431 CysArgAlaProAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 450
QY 1261 GACTGACCGAGCGCGCGCGCGCACTTCTGCGCGAGATGCTGCGCGCGCGCGCGCGCGCG 1320
    |||
Db 451 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 470
QY 1321 CCGCGCGCACTTCTCTGCAAGAACCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 1380

```

```

Db 471 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 484
QY 1381 CCGCGCGAGAGTTCGCTTCGAGAGACACACCGCGCGCGCGCGCGCGCGCGCGCGAGAC 1440
    |||
Db 485 ProAlaGlnSerPheArgPheGluGluTrpThrProAlaProLysGlnGluProIleGlu 504
QY 1441 CCGGAGCCCTTACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1500
    |||
Db 505 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 520
QY 1501 AGCCAG 1506
    |||
Db 521 SerGln 522

RESULT 7
ADCT2879
ID ADCT2879 standard; protein; 492 AA.
XX
AC ADCT2879;
XX
DT 18-DEC-2003 (first entry)
XX
DE HIV-1 gag protein containing an NCP sequence, SEQ ID NO 109.
XX
KW complex; HIV nucleocapsid protein 7; NCP7; HIV-ps1-site; anti-HIV;
KW human immunodeficiency virus; HIV.
XX
OS Human immunodeficiency virus 1.
XX
PN WO2003060098-A2.
XX
PD 24-JUL-2003.
XX
PF 10-JAN-2003; 2003WO-US000801.
XX
PR 11-JAN-2002; 2002US-0347369P.
XX
PA (ACHI-) ACHILLION PHARM INC.
XX
PI Beutcher D, Hou X, Marlor CW, Rice WG, Yang W;
XX
DR WPI; 2003-646042/61.
XX
PT Method of determining whether a compound inhibits formation of complex
PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site
PT oligonucleotide by comparing amount of complex formed in presence/absence
PT of compound.
XX
PS Claim 2; SEQ ID NO 109; 105bp; English.
XX
CC The invention relates to a novel method of determining whether a compound
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)
CC polypeptide and an HIV-ps1-site oligonucleotide. The method involves
CC adding the HIV-ps1-site oligo to a mixture of the NCP7 polypeptide and
CC the novel compound and comparing amount of complex formed, with that
CC formed in the absence of the compound, where a decrease in the amount of
CC complex formed in presence of the compound indicates that the compound
CC inhibits complex formation. The method of the invention involves anti-HIV
CC activity. The compounds identified are useful for treating a subject
CC infected with human immunodeficiency virus (HIV) by administering the
CC compound to the subject. This sequence represents an HIV-1 gag protein
CC which contains an NCP sequence of the invention.
XX
SQ Sequence 492 AA;
XX

Alignment Scores:
Pred. No.: 6,79e-131 Length: 492
Score: 2440.00 Matches: 464
Percent Similarity: 94.42% Conservative: 10
Best Local Similarity: 92.43% Mismatches: 18
Query Match: 85.92% Indels: 10
DB: Gaps: 2

```



US-09-475-704a-4 (1-1509) x ADC72879 (1-492)

QY 1 ATGGGCGCCGGCCGACATCTCGCGGGGAGAGCTGGACAAGTGGAGAAATCCGC 60  
 DB 1 MetGlyAlaSerAlaSerIleLeuArgGlyGlyLeuAspIleTrpGlyIleArg 20  
 QY 61 CTGGCGCCGGGGGCAAGAACCTACATGCTGAACCACTGGTGGGCCCGCGCGAG 120  
 DB 21 LeuArgProGlyGlyLeuValyIleValyLeuValIleValAlaSerArgGln 40  
 QY 121 CTGGAGGGCTTCGCTCGAACCCTGCTGGAGACCGCGAGGGCTGCAAGCATC 180  
 DB 41 LeuGluArgPheAlaLeuAsnSerGlyLeuLeuGluThrAlaGluIleValyIle 60  
 QY 181 ATGAAGACCTGCAGCCGCGCTGCAAGCCGCGACCGAGAGCTGCGACCTGTACAC 240  
 DB 61 IleValGlnLeuGlnProAlaLeuGlnThrGlyThrGluIleuValySerLeuVal 80  
 QY 241 ACCGTGGCCACCTGTATCTGCGTGGAGCGCGGATGAGTCCGCGACCAAGAGGCG 300  
 DB 81 ThrValAlaThrLeuValCysValHisAlaGlyIleGluValAlaArgAspThrIleVal 100  
 QY 301 CTGCAACAAGATCGAGGAGGAGCAACAAGTCCGACGACAGAAAGCCAGAGCGAGAG 360  
 DB 101 LeuAspValyIleGluIleuGlnIleuValyIleValyIleValyIleValyIle 120  
 QY 361 GCGGACGCGAAGGTGAGCCAGAACTACCCATCTGTGACAGAACTGAGGCGCATGTGT 420  
 DB 121 AlaAspGlyIleValSerGlnIleuValProIleValGlnAsnLeuGlnIleVal 140  
 QY 421 CACCAAGGCATTCAGCCCGCGACCCCTGAAGCGCTGGGTGAAGTATCGAAGAAAGGCG 480  
 DB 141 HisGlnProIleSerProArgThrLeuAsnAlaIleValIleGluIleuValyAla 160  
 QY 481 TTCAGCGCCGAGGTGATCCCATGTTCAACGCGCTGAGGAGGAGGCGACCCCGCGAGAC 540  
 DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluIleValAlaThrProGlnAsp 180  
 QY 541 CTGAACACGATGTTGAACAACCGTGGGCGGACACGAGCGCGCCATGCAATGCTGAAGAC 600  
 DB 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuValyAsp 200  
 QY 601 ACCATCAACAGAGAGCGCGCGAGTGGAGACCGCTGCAACCCGCTGAGAGCGCGCGCGCTG 660  
 DB 201 ThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisIleValyIleProVal 220  
 QY 661 GCGCCGCGGAGATGCGGACGCGCGCGGAGGAGCATGCGCGCGACGACGACCGCTG 720  
 DB 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
 QY 721 CAGGAGCAGATGCGCTGATGACACGAAACCCCGCGTGGCGCGGAGCATTTACAG 780  
 DB 241 GlnGluGlnIleThrThrMetThrSerAsnProProValProValGlyAspIleValy 260  
 QY 781 CCGTGATATCTCTGGGCTGAAACAAGATCGTGGATGTACAGCCCGCTGAGCATCTCTG 840  
 DB 261 ArgTrpIleIleLeuGlyLeuAsnValyIleValArgMetIleSerProValSerIleLeu 280  
 QY 841 GACATCCGCGGAGGCGCGCAAGAGACCTTCGCGGATAGTGGACCGCTTCTTCAAGAC 900  
 DB 281 AspIleValyGlnGlyProValyIleProValyIleProValyIleValyIleValy 300  
 QY 901 CTGGCGCGCGAGCAGGCGCAACGAGACGTAAGAACTGTAGTACGAGACCTGTGCTGTG 960  
 DB 301 LeuArgAlaGluGlnIleThrGlnAspValyIleValyIleValyIleValyIleVal 320  
 QY 961 CAGAACGCCAACCCCGATCTGCAAGACCATCTCGCGCTCTGGGCGCGCGCGACCTGTG 1020  
 DB 321 GlnAsnAlaAsnProAspCysIleThrIleLeuArgAlaLeuGlyProAlaAlaSerIleu 340  
 QY 1021 GAGGAGATGATGACCGCTGCGCGGCGTGGGCGCGCGCGCGCAAGAGCGCGCGCTG 1080

DB 341 GlnIleMetMetThrAlaCysGlnGlyValGlyGlyProSerHisIleValAlaArgValLeu 360  
 QY 1081 GCGGAGCGGATGAGCCAGGCGCAACGCGTGAACATCATGATTCAGAAAGCAACTTCAAG 1140  
 DB 361 AlaGluAlaMetSerGlnAlaAsnThrThrAsnIleMetMetGlnIleSerAsnPheValy 380  
 QY 1141 GCGCCCGCGCGCAAGCTCAAGTCTTCAACTCGGCAAGAGGCGCACTCCGCAAGAC 1200  
 DB 381 GlyProArgArgThrValIleCysPheAsnValyIleValyIleValyIleValyIle 400  
 QY 1201 TGCAGCGCGCGCGCAAGAGGCTGTGAAGTGTGCGGCAAGAGCGCCAGCATGAG 1260  
 DB 401 CysArgAlaProArgIleValyIleCysIleValyIleValyIleValyIleValyIle 420  
 QY 1261 GACTGCAACGAGCGCGCAAGCTTCTGGGCAAGATGTGGCCAGCCAGCAAGGGCGCG 1320  
 DB 421 AspCysIleGluArgGlnAlaAsnPheLeuGlyIleValyIleTrpProSerHisIleValy 440  
 QY 1321 CCGGCAACTTCTGTGAGAACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
 DB 441 ProGlyAsnPheLeuGlnIleSerArgProGlu-----ProThrAlaPro 454  
 QY 1381 CCGCGCGAGAGCTTCCGCTTCGAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCG 1440  
 DB 455 ProAlaGluSerPheArgPheGluGluThrThrProAlaProIleGluIleProIleValy 474  
 QY 1441 CCGGAGCCCTTACCGGAGCGCGCTGACCGCGCTGCGGAGCTTGTGGCGAGCGCGCGCTG 1500  
 DB 475 -----ArgIleProLeuThrSerLeuValySerLeuPheGlySerAspProLeu 490  
 QY 1501 AGCGAG 1506  
 DB 491 SerGln 492

RESULT 8  
 ID ADC72880  
 AD ADC72880 standard, protein; 492 AA.  
 AC ADC72880;  
 DT 18-DEC-2003 (first entry)  
 DE HIV-1 gag protein containing an NCP sequence, SEQ ID No 110.  
 KW complex; HIV nucleocapsid protein 7; NCP7; HIV-ps1-site; anti-HIV;  
 KW human immunodeficiency virus; HIV.  
 OS Human immunodeficiency virus 1.  
 PN WO2003060098-A2.  
 PD 24-JUL-2003.  
 PF 10-JAN-2003; 2003WO-US000801.  
 PR 11-JAN-2002; 2002US-0347369P.  
 PA (AChI-) ACHILION PHARM INC.  
 PI Beutler D, Hou X, Marlor CW, Rice WG, Yang W;  
 DR WPI; 2003-646042/61.  
 PT Method of determining whether a compound inhibits formation of complex  
 PT between HIV nucleocapsid protein 7 polypeptide and HIV ps1-site  
 PT oligonucleotide by comparing amount of complex formed in presence/absence  
 PT of compound.  
 PS Claim 2; SEQ ID NO 110; 105bp; English.  
 CC The invention relates to a novel method of determining whether a compound  
 CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)  
 CC polypeptide and an HIV-ps1-site oligonucleotide. The method involves

CC adding the HIV-psi-site oligo to a mixture of the NCP7 polypeptide and  
 CC the novel compound and comparing amount of complex formed, with that  
 CC formed in the absence of the compound, where a decrease in the amount of  
 CC complex formed in presence of the compound indicates that the compound  
 CC inhibits complex formation. The method of the invention involves anti-HIV  
 CC activity. The compounds identified are useful for treating a subject  
 CC infected with human immunodeficiency virus (HIV) by administering the  
 CC compound to the subject. This sequence represents an HIV-1 gag protein  
 CC which contains an NCP sequence of the invention.

XX Sequence 492 AA;

Alignment Scores:

Pred. No.:	2,52e-130	Length:	492
Score:	2430.00	Matches:	458
Percent Similarity:	94.82%	Conservative:	18
Best Local Similarity:	91.24%	Mismatches:	16
Query Match:	85.56%	Indels:	10
DB:	7	Gaps:	2

US-09-475-704A-4 (1-1509) x ADC72880 (1-492)

```

QY 1 ATGGGCGCCCGCCGAGCATCTGCGCGGAGAAAGTGGACAGTGGAGAAATCCGC 60
DB 1 MetGlyAlaArgAlaSerLleLeuArgGlyGlyLeuAspThrTrpGluLysIleArg 20
QY 61 CTGGCGCCCGCGGAGAAAGCATCACTGAGAGCACTGCTGGGCGCCGCGGAG 120
DB 21 LeuArgProGlyGlyLysLysCysTrpMetLeuLysIleValTrpAlaSerArgGlu 40
QY 121 CTGAGAGGCTTCCGCTGAAACCCCGGCTGTGTGAGAGCCGCGAGGGCTGCAAGCATC 180
DB 41 LeuGluArgPheSerLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
QY 181 ATGAAGCAGCTCTCAACCGCCCTGTGACAGCCGAGACCTGCGAGCTGTACAC 240
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyTrpGluLysSerLeuLysAsn 80
QY 241 ACCGTGGCACCCTGTACTGTCGTCGACGCGCGCATCGAGTCCGCGACCAAGAGGCC 300
DB 81 ThrValAlaThrLeuPheCysValHisGluLysIleAlaValArgAspThrLysGluAla 100
QY 301 CTGACCAAGATCGAGAGAGAGCAACAAGTCCAGCAGAGAACCCAGAGGCCAGAG 360
DB 101 LeuAspLysIleGluGluGlnGlnAsnLysSerGlnGlnLysThrGlnAlaLysAla 120
QY 361 GCGGACGCGCAGAGTGAAGCAAGATCACTGTCGAGAACCTGACAGGCGCAGATG 420
DB 121 AlaAspGlyThrValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlyGlnMetVal 140
QY 421 CACGAGCCATCAGCCCGCAGCACTGTAAGCCTGGTGAAGTATGAGAGAGAGGCC 480
DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLysAla 160
QY 481 TTCAGCCCGGAGGTATCCCATGTTCAACCGCTGAGCAGAGGGGCCACCCCGAGAC 540
DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGluAsp 180
QY 541 CTGAACAGATGTTGAACCGGTGGCGGCGCACCGGCGCCATGCAATGCTGAAGAC 600
DB 181 LeuAsnTrpMetLeuAsnThrValGlyGlnHisGlnAlaAlaMetGlnMetLeuLysAsp 200
QY 601 ACCATCAAGAGAGCGCGCAGAGTGAAGCAGCTGACCCCGTGCAGGCGCGCCGCTG 660
DB 201 ThrIleAsnGluGluAlaHisGluTrpAspArgLeuHisProValHisAlaGlyProIle 220
QY 661 GCGCCCGGCGCAGATCGGACCCCGCGCGCAGCGACATGCGCGGCGCACAGCACCTG 720
DB 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240
QY 721 CAGAGAGCAGATCGCTGATGACCAAGCAACCCCGCTGCGCGGCGGAGATCTACAG 780
DB 241 GlnGluGlnIleValTrpMetThrAsnAsnProProValProValGlyAspIleLysIle 260

```

```

QY 761 CGGTGATCATCTGGGCTTGAAACAAGTCTGTGGGATGTACAGCCCGCTGACATCTG 840
DB 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValAlaArgMetLysSerProValSerIleLeu 280
QY 841 GACATCGCGCAGGGGCGCCAGAGAGCCCTTCCGCGCATCACTGCGGAGCCGCTTCAAGAC 900
DB 281 AspIleArgGlnGlyProLysGluProPheArgAspIleValAspArgPheLysThr 300
QY 901 CTGGCGCGCAGAGCGCCACAGACGCTGAAGAAGTGGATGACGAGACCCCTGTGTG 960
DB 301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320
QY 961 CAGAAAGCGCAACCCCGATGGAAGCATCTCTGGCGCTTCCGCGCCCGCGGCGCACCTG 1020
DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu 340
QY 1021 GAGGAGATGATGACCGCTGCGCAGGCGCTGGCGCGCCCGCGCCAAAGCGCGGTGCTG 1080
DB 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysValAlaValLeu 360
QY 1081 GCCAGAGCATGAGCCAGCCAGCCAGCAGCGTGAACATCATGATGACAAAGCAACTTCAAG 1140
DB 361 AlaGluAlaMetSerGlnThrAsnAsnAlaAsnIleMetMetGlnArgSerAsnPheLys 380
QY 1141 GCGCCCGCGCGCAACGTCAATGCTTCACTGCGCGCAAGAGGGCGCCATGCGCCAAAGAC 1200
DB 381 GlyProArgArgIleIleLysCysPheAsnLysCysGlyLysGluGlnLysLeuAlaArgAsn 400
QY 1201 TGCGCGCGCCCGCCAGAAAGGCTGTGAAGTGGCGGCAAGAGAGGCCACCAATGAAG 1260
DB 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnLysGlnMetLys 420
QY 1261 GACTGCAACGAGCGCCAGGCCAATCTCTGGGCAAGATCTGGCCAGCCACCAAGGCGCC 1320
DB 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGCGCAACTCTCTGCAAGAACCGGAGGAGCGCGCGCCCGCCACCGGCCACCGCGCCC 1380
DB 441 ProLysAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454
QY 1381 CCGCGCAGAGCTTCCGCTTGAAGAGACCAACCCCGCCCGCCAGCAGAGCCCAAGAC 1440
DB 455 ProAlaGluSerPheArgPheGlnGluThrThrProAlaProLysGlnLysProArgGlu 474
QY 1441 CGCAGGCCCTAACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTGCGCAGCGCGCCCTG 1500
DB 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490
QY 1501 AGCCAG 1506
DB 491 SerGln 492

```

RESULT 9  
 ID ADC72910 standard; protein; 491 AA.  
 XX ID ADC72910;  
 AC ADC72910;  
 XX 18-DEC-2003 (first entry)  
 DT  
 XX HIV-1 gag protein containing an Ncp sequence, SEQ ID No 140.  
 DE complex; HIV nucleocapsid protein 7; NCP7; HIV-psi-site; anti-HIV;  
 KW human immunodeficiency virus; HIV.  
 XX Human immunodeficiency virus 1.  
 OS  
 XX MO2003060098-A2.  
 XX 24-JUL-2003.  
 PD 10-JAN-2003; 2003MO-US000801.  
 PF

XX 11-JAN-2002; 2002US-0347369P.  
 PR (ACHT-) ACHILLION PHARM INC.  
 PA Beuchter D, Hou X, Marlor CM, Rice WG, Yang W;  
 XX WPI; 2003-646042/61.  
 XX  
 XX Method of determining whether a compound inhibits formation of complex  
 PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site  
 PT oligonucleotide by comparing amount of complex formed in presence/absence  
 of compound.  
 XX  
 XX Claim 2; SEQ ID NO 140; 105bp; English.  
 XX  
 XX The invention relates to a novel method of determining whether a compound  
 CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)  
 CC polypeptide and an HIV-psi-site oligonucleotide. The method involves  
 CC adding the HIV-psi-site oligo to a mixture of the NCP7 polypeptide and  
 CC the novel compound and comparing amount of complex formed, with that  
 CC formed in the absence of the compound, where a decrease in the amount of  
 CC complex formed in presence of the compound indicates that the compound  
 CC inhibits complex formation. The method of the invention involves anti-HIV  
 CC activity. The compounds identified are useful for treating a subject  
 CC infected with human immunodeficiency virus (HIV) by administering the  
 CC compound to the subject. This sequence represents an HIV-1 gag protein  
 CC which contains an NCP sequence of the invention.  
 CC  
 SQ Sequence 491 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 7,68e-130 Length: 491  
 Score: 2421.50 Matches: 461  
 Percent Similarity: 94.82% Conservative: 15  
 Best Local Similarity: 91.83% Mismatches: 11  
 Query Match: 85.26% Indels: 11  
 DB: 7 Gaps: 3  
 US-09-475-704A-4 (1-1509) x ADC72910 (1-491)

QY 481 TTCAGCCCGAGGTGATCCCATGTTACCGCCCTGAGAGGAGGCGCACCCCGCAGAGC 540  
 |||||  
 DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerIleuValIleThrProGluAsp 180  
 |||||  
 QY 541 CTGAACACGATGTTGAACACCGTGGCGGCGCACAGCGCCATGACATGCTGAAGAC 600  
 |||||  
 DB 181 LeuAenThMetLeuAenThValIleGlyIleGlnAlaIleMetGlnMetLeuIleAsp 200  
 |||||  
 QY 601 ACCATCAACGAGAGGCGCCGAGTGGACCGCTTCGACCCCTGACGCGCGCCCTG 660  
 |||||  
 DB 201 ThrIleAsnGlnGluIleAlaIleGluThrPheAspArgLeuIleProValIleAlaGlyProIle 220  
 |||||  
 QY 661 GCCCGCGGCGAGATGGCGACCGCGGCGAGGACATGCGCGCGCGCACAGACCCCTG 720  
 |||||  
 DB 221 AlaProGlyIleMetArgIleProArgIleProArgIleAspIleAlaGlyThrThrSerSerLeu 240  
 |||||  
 QY 721 CAGAGACATGCGCTGATGACACAGAACACCCCGCTGCGCCCTGGGCGCATCTACAG 780  
 |||||  
 DB 241 GlnGlnGlnIleAlaIleThrMetThrGlyAsnProProValProValGlyAspIleTyrLe 260  
 |||||  
 QY 781 CGGTGATCATCTGGGCTTGAACAAAGATCGTGGATGATACAGCCCGTGAATCTG 840  
 |||||  
 DB 261 ArgTrpIleIleLeuGlyLeuAsnIleValAlaArgMetTyrSerProValSerIleLeu 280  
 |||||  
 QY 841 GACATCGCGCAGGCGCCCAAGAGCGCTTCGCGACTAGTGGACCGCTTCAAGACC 900  
 |||||  
 DB 281 AspIleLeuGlnGlyProIleGluProPheArgAspTyrValAspArgPheIleIleTyr 300  
 |||||  
 QY 901 CTGCGCGCGAGCAGGCGCACCGAGACGTGAAGAACTGTGATGACCGACCTTGCTGCTG 960  
 |||||  
 DB 301 LeuArgAlaGlnGlnIleAlaIleThrMetThrGlyAsnProValValAsnTrpMetThrPheLeuVal 320  
 |||||  
 QY 961 CAGAACGCCAACCCCGCATGCAAGACATCTGCGGCTTGGCGCCCGCGCACCTG 1020  
 |||||  
 DB 321 GlnAsnAlaAsnProAspCysIleThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
 |||||  
 QY 1021 GAGGAGATGATACCGCTGCGACGCGCGGCGCGCGCGCGCGCACAGGCGCGCTG 1080  
 |||||  
 DB 341 GlnGlnMetMetThrIleCysGlnGlyValIleGlyIleProSerIleIleValAsnValLeu 360  
 |||||  
 QY 1081 GCCGAGCGCATGAGCGCGCACCGCATGCAAGCTGATGATGACGAGAGCAACTTCAAG 1140  
 |||||  
 DB 361 AlaGlnAlaMetSerIleThrAsnSer---ThrIleLeuMetGlnArgSerAsnPheIle 379  
 |||||  
 QY 1141 GGCCTCCGCGCAACGCTCAAGTCTTCAACTGCGGCGAGAGAGGCGCACTCGCAAGAA 1200  
 |||||  
 DB 380 GlyProIleArgIleValIleCysPheAsnIleGlyIleGlnGlyIleIleAlaIleAsn 399  
 |||||  
 QY 1201 TGCCTGCGCGCGCGCAAGAGGCGCTGCGAAGTGGCGGAGAGGCGCACAGATGAAG 1260  
 |||||  
 DB 400 CysArgAlaProArgIleIleValIleCysIleIleValIleGlnGlyIleIleMetIle 419  
 |||||  
 QY 1261 GACTGACCGAGCGCGCAAGCACTTCTGGGCAAGATCTGCGCCCGCACAGAGGCGCG 1320  
 |||||  
 DB 420 AspCysThrGlnArgGlnAlaAsnPheLeuGlyIleIleTrpProSerIleIleValIle 439  
 |||||  
 QY 1321 CCGCGCACTTCTGCAAGAACCGCAGAGCGCGCGCGCGCGCACCGTGCACCGCGCCC 1380  
 |||||  
 DB 440 ProGlyAsnPheLeuGlnIleSerArgProGlu-----ProThrAlaPro 453  
 |||||  
 QY 1381 CCGCGCGAGAGCTTCCGCTTGAAGAGACACCCCGCGCGCGCGCAAGAGAGCGCAAGAG 1440  
 |||||  
 DB 454 ProAlaGlnSerPheArgPheGlnGlnIleThrThrProAlaLeuGlnIleIleProIle 473  
 |||||  
 QY 1441 CCGGAGCCCTTACCGGAGCGCGCTGACCGCGCTGCGCAGCGCTTGGCGAGCGCGCCCTG 1500  
 |||||  
 DB 474 -----ArgIleProLeuThrSerIleLeuArgSerIlePheGlySerAspProLeu 489  
 |||||  
 QY 1501 AGCCAG 1506  
 |||||  
 DB 490 SerGln 491  
 |||||

RESULT 10



Db 477 ProAlaGluSerPheGluGluGluThrThrProAlaProLysGlnGluProLysAsp 496  
QY 1441 CGGAGACCTTACCGCGAGCCCTTGACGCGCTGTCGCGACGCTTTCGCGAGCGGCCCTG 1500  
Db 497 -----ArgLysProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 512  
QY 1501 AGCGAG 1506  
Db 513 SerGln 514  
RESULT 11  
ADC72913  
ID ADC72913 standard; protein; 491 AA.  
XX  
AC ADC72913;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE HIV-1 gag protein containing an NCP sequence, SEQ ID NO 143.  
XX  
KM complex; HIV nucleocapsid protein 7; NCP7, HIV-psi-site; anti-HIV;  
KM human immunodeficiency virus; HIV.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN MO2003060098-A2.  
XX  
PD 24-JUL-2003.  
XX  
PF 10-JAN-2003; 2003WO-US000801.  
XX  
PR 11-JAN-2002; 2002US-0347369P.  
XX  
PA (ACHI-) ACHILLION PHARM INC.  
XX  
PI Beutler D, Hou X, Marlor CM, Rice WG, Yang W;  
XX  
DR WPI; 2003-646042/61.  
XX  
PT Method of determining whether a compound inhibits formation of complex  
PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site  
PT oligonucleotide by comparing amount of complex formed in presence/absence  
PT of compound.  
XX  
PS Claim 2; SEQ ID NO 143; 105BP; English.  
XX  
XX The invention relates to a novel method of determining whether a compound  
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)  
CC polypeptide and an HIV-psi-site oligonucleotide. The method involves  
CC adding the HIV-psi-site oligo to a mixture of the NCP7 polypeptide and  
CC the novel compound and comparing amount of complex formed, with that  
CC formed in the absence of the compound, where a decrease in the amount of  
CC complex formed in presence of the compound indicates that the compound  
CC inhibits complex formation. The method of the invention involves anti-HIV  
CC activity. The compounds identified are useful for treating a subject  
CC infected with human immunodeficiency virus (HIV) by administering the  
CC compound to the subject. This sequence represents an HIV-1 gag protein  
CC which contains an NCP sequence of the invention.  
XX  
SQ Sequence 491 AA;  
Alignment Scores:  
Pred. No.: 2,19e-129 Length: 491  
Score: 2413.50 Matches: 456  
Percent Similarity: 95.02% Conservative: 21  
Best Local Similarity: 90.84 Mismatches: 14  
Query Match: 84.98 Indels: 11  
DB: 7 Gaps: 3  
US-09-475-704A-4 (1-1509) x ADC72913 (1-491)  
QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGAGAGCTGAGCAAGTGGAGAAATCCGC 60  
|||||

Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGlyLysIleArg 20  
QY 61 CTGGCGCCCGGGGCGCAAGAAAGCATACATGCTGAGACACCTGTGTGGCCAGCCCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisTrpMetIleLysIleLeuValTrpAlaSerArgLys 40  
QY 121 CTGGAGGGCTTGCCCGCTGAACCCCGGCGCTGGAGAACCGCGAGGGCTGCAAGCATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlyLysGlnIle 60  
QY 181 ATGAAGACCTGACAGCCCGCTGACAGCCGACCGACCGAGAGCTGGACCTGTACAC 240  
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLeuArgSerLeuHisAsn 80  
QY 241 ACCGTGGCCACCTGTGATGCTGCTGCAAGCGCGCATGAGCTCCGACACCAAGAGCC 300  
Db 81 ThrValAlaThrLeuLysCysValHisValGlyIleGluIleArgSerThrLysGlnAla 100  
QY 301 CTGGAACAAGATGAGAGGAGCAAGAAAGTCCAGACAGACCAAGACCGAGCGCAAGAG 360  
Db 101 LeuAspLysIleGluGluGluGlnGluLysSerGlnGlnLysThrGlnIleAlaLysGlu 120  
QY 361 GCCGACGCGCAAGTGAAGCCAGAACTACCCCATCTGTGACAGACCTGACAGCGCGAGATG 420  
Db 121 AlaAspGlyLysValSerGlnAsnTrpProIleValGlnAsnLeuGlnIleGlnMetVal 140  
QY 421 CACCAAGGCATAGAGCCCGCGACCCCTGAAGCGCTGTGAGAGTGTGAGAGAGGCC 480  
Db 141 HisGlnAlaLeuSerProArgThrLeuAsnAlaTrpValIleGluGluLysAla 160  
QY 481 TTCAGCCCGGAGGTATATCCCATGTTCAACGGCGCTGAGAGAGGCGCGACCCCGAGAG 540  
Db 161 PheSerProGlnIleIleProMetPheThrAlaLeuSerGlnLysAlaThrProGlnAsp 180  
QY 541 CTGAACACGATGTGACACCGTGGCGGCGCACAGCGCCCATGACGATGTGAAAGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCGCTTGACCCCGTGCAGGCGCGCCGCTG 660  
Db 201 ThrIleAsnGluGlnAlaIleAlaIleTrpAspArgLeuHisProAlaGlnIleArgProIle 220  
QY 661 GCGCCCGCGAGATGCGGACCGCGCGGACGACATGCGCGCGCGCCACCGACCGCCG 720  
Db 221 AlaProGlyGlnMetArgLysProArgLysSerAspIleAlaGlyThrThrSerThrLeu 240  
QY 721 CAGAGACGATGCGCTGATGACACCAAGACCCCGCTGCGCGTGGGCGACATTTACAAG 780  
Db 241 GlnGluGlnIleAlaIleTrpMetThrGlyAsnProProValProValGlyLysIleTrpLys 260  
QY 781 CGTGGATCATCTGGGCGCTGAAACAAGATGTCGAGATGTACAGCCCGCTGAGCATCTG 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIleLeu 280  
QY 841 GACATCCGCGAGGCGCGCAAGAGACCTTCGCGCATAGTGGACCGCTTTCTTAAGAC 900  
Db 281 AspIleLysGlnGlyProLysGlnProPheArgAspTrpValAspArgPheHeuLysThr 300  
QY 901 CTGGCGCGCGAGACGCGCAAGACGATGAGAGAGTGAAGTGAATGACGAGACCTGTGCTG 960  
Db 301 LeuArgAlaGluGlnAlaIleThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320  
QY 961 CAGAACGCGCAACCGCGATGCAAGACATCTGCGCGCTTGGGCGCGCGCCACCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
QY 1021 CAGGAGAGTGAAGCGCGCTGCGAGGCGTGGCGCGCGCGCGCAAGAGCGCGCTGCTG 1080  
Db 341 GlnGluMetMetThrAlaCysGlnIleValGlyGlyProSerHisLysValaArgValLeu 360  
QY 1081 GCCGAGGCGATGAGCGAGCGCAAGCGCTGAACATCATGATGACAGAGCAACTTCAAG 1140  
Db 361 AlaGluAlaMetSerGlnThrAsnAsn---SerIleLeuMetGlnArgSerAsnHisLys 379

QY 1141 GGCCCCCGGCGCAAGTCAAGTCTTCAACTGGCGCAAGAGGCGCCATCGCCCAAGAAC 1200  
Db 380 GlyPheLysEaRgThrValLysCysPheAsnCySGlyLysGlnGlyHisIleAlaAsn 399  
QY 1201 TGGCGGCGCCCGCGCAAGAGGCTGCTGGAGTGGCGCAAGAGGCGCCATGATGAAG 1260  
Db 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisIleMetLys 419  
QY 1261 GACTGCAACGAGGCGCCCAAGCTTCTGGGCAAGATTTGGCGCCCAAGAGGCGCGC 1320  
Db 420 AspCysThrGlnArgGlnAlaAsnPheLysGlyLysIleTrpProSerHisLysGlyArg 439  
QY 1321 CCGCGCAACTCTTGGAGAACGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
Db 440 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 453  
QY 1381 CCGCGCGAGAGCTTCCGCTTTCAGAGAGACACCCCGCGCGCGCGCGCGCGCGCGCG 1440  
Db 454 ProAlaGlnSerPheArgPheGlnGlnTrpThrProAlaLeuLysGlnGlnLysAsp 473  
QY 1441 CCGGAGCCCTACCGCGCAAGCCCTGACCGCCCTGCGAGCTGTTCGGCAGCGCGCGCTG 1500  
Db 474 -----ArgGlnProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 489  
QY 1501 AGCCAG 1506  
Db 490 SerGln 491  
RESULT 12  
ID ADN36396  
ID ADN36396 standard; protein; 514 AA.  
XX  
AC ADN36396;  
XX  
DT 15-JUL-2004 (first entry)  
XX  
DE Human protein for anti-HIV vaccine.  
XX  
KM anti-HIV; vaccine; HIV; promoter; viral particle; immunization.  
OS Homo sapiens.  
XX  
PN WO2004035006-A2.  
XX  
PD 29-APR-2004.  
XX  
PF 17-OCT-2003; 2003WO-US033112.  
XX  
PR 18-OCT-2002; 2002US-0419465P.  
XX  
PA (AARO-) AARON DIAMOND AIDS RES CENT.  
XX  
PI Huang Y, Ho DD, Chen Z;  
XX  
DR WPI: 2004-348328/32.  
DR N-PSDB; ADN36395.  
XX  
Nucleic acid vector comprising at least one HIV sequence operably linked  
PT to a promoter and encoding a protein that does not assemble into viral  
PT particles, useful in immunizing a subject against HIV infection.  
XX  
PS Disclosure; SEQ ID NO 10; 16pp; English.  
XX  
CC The invention relates to a nucleic acid vector comprising at least one  
CC HIV sequence operably linked to a promoter and encoding a protein that  
CC does not assemble into viral particles. The nucleic acid vector is useful  
CC in immunizing a subject against HIV infection. This sequence corresponds  
CC to a peptide used in the invention.  
XX  
SQ Sequence 514 AA;  
Alignment Scores:

Prod. No.: 3,266-129  
Score: 2410.50  
Percent Similarity: 94.82%  
Best Local Similarity: 91.04%  
Query Match: 84.88%  
DB: 8 Gaps: 3  
US-09-475-704A-4 (1-1509) x ADN36396 (1-514)  
QY 1 ATGGGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 24 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGlnLysIleArg 43  
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 44 LeuArgProGlyGlyLysLysHisIleTrpMetLeuLysHisIleValTrpAlaSerArgGlu 63  
QY 121 CTGAGAGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
Db 64 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnTrpSerGlnGlyCysLysGlnIle 83  
QY 181 ATGAAGAGCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 84 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLysLeuArgSerLeuPheAsn 103  
QY 241 ACCGTGGCCACCTGTACTGCTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 104 ThrValAlaTrpLeuLysCysValHisGlnGlyLysIleGlnIleArgAspThrLysGlnAla 123  
QY 301 CTGCAAGATCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360  
Db 124 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 143  
QY 361 GCCGAGCGCAAGGTGAGCGCAAGCTACCCATCGTGCAGAACTTCGAGGCGCGCGAGGTG 420  
Db 144 AlaAspGlnLysValSerGlnAsnLysProIleValGlnAsnLeuGlnGlnIleMetVal 163  
QY 421 CACGAGCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Db 164 HisGlnAlaIleSerProAlaGlnLeuAsnAlaTrpValLysValIleGlnGlnLysAla 183  
QY 481 TTCAGCGCGCGAGTATCCCATGTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
Db 184 PheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsp 203  
QY 541 CTGAACACAGATGTTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
Db 204 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLysAsp 223  
QY 601 ACCATCAACGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
Db 224 ThrIleAsnGlnGlnAlaIleGlnTrpAspArgValHisProValHisIleAlaGlyProIle 243  
QY 661 GCCCGCGCGAGATGCGCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
Db 244 AlaProGlyGlnMetArgLysProArgGlySerAspIleAlaGlyThrTrpSerThrLeu 263  
QY 721 CAGAGAGAGATCGCGCTGGATGACACGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
Db 264 GlnIleGlnIleAlaTrpMetThrSerAsnProProValProValGlyGlnIleLys 283  
QY 781 CGGTGATCATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840  
Db 284 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIleLeu 303  
QY 841 GACTCGCGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
Db 304 AspIleLysGlnGlyProLysGlnProPheArgAspLysValAspArgPhePheLysThr 323  
QY 901 CTGCGCGCGAGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
Db 324 LeuArgAlaGlnGlnAlaTrpGlnAspValLysAsnTrpMetThrAspThrLeuVal 343

```
QY 961 CAGAACGCCAACCCGCTGCAAGACCATCTGCGGCGTCTGGCCCGGCGCCAGCCCTG 1020
   |||||
Db 344 GlnMetAlaAsnProAspCysIleValIleuArgAlaLeuGlyProGlyAlaSerLeu 363
QY 1021 GAGGAGATATGACCGCTGCGCAGAGGCGTGGGCGCCCGCCAGAGCCCGCTGCTG 1080
   |||||
Db 364 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisValAlaArgValLeu 383
QY 1081 GCCGAGCCATATGACCGCCAGCCAGACCGTGAACATCATATGATGAGAGAGCAATCTTCAAG 1140
   |||||
Db 384 AlaGluAlaMetSerGlnAlaAsnGly---ThrIleLeuMetGlnArgSerAsnPheLeu 402
QY 1141 GACCCCGCGCGCAACGTCAAGTGTCTCAACTGCGGAGAGGAGGCGCAATCGCCAGAGAC 1200
   |||||
Db 403 GlySerIleValArgIleValIleValCysPheAsnGlyLeuGlnGlyHisIleAlaArgAsn 422
QY 1201 TSCCGGCGCCCGCCAGAGAGGCGTCTGGAGATGCGGAGAGGAGGCGCAAGATGAAG 1260
   |||||
Db 423 CysArgAlaProArgIleValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 442
QY 1261 GACTGCAACGAGCGCGCCAGCCCACTTCTGGGCGCAATGTGCGCCAGCCAGAGGCGCCG 1320
   |||||
Db 443 AspCysThrGluArgGlnAlaAsnPheLeuGlyLeuIleTrpProSerHisValArg 462
QY 1321 CCCGCGCACTTCTGCGAGAACCGCGAGAGCGCGCGCCCGCCAGCTGCGCCAGCCCGCC 1380
   |||||
Db 463 ProGlyAsnPheLeuIleAsnProGlu-----ProThrAlaPro 476
QY 1381 CCCGCGAGAGCTTCTGCTTCCAGAGAGACACCCCGCCCGCCAGAGAGGCGCCAGAGAC 1440
   |||||
Db 477 ProAlaGluSerPheArgPheGlnGluIleThrProAlaProGlnGlnIleProGlyAsn 496
QY 1441 CCCGAGCCCTTACCGCGAGCGCCCTGACCGCCCTGCGAGCGCTGTTCGCGAGCGCCCTG 1500
   |||||
Db 497 -----ArgIleProLeuThrSerLeuIleValSerLeuPheGlySerAsnProLeu 512
QY 1501 AGCCAG 1506
   |||||
Db 513 SerGln 514

RESULT 13
AAB69276
ID AAB69276 standard; protein; 498 AA.
XX
AC AAB69276;
XX
DT 12-SEP-2003 (revised)
DT 20-APR-2001 (first entry)
XX
DE HIV-1 non-subtype B clone 962M751-3 gag protein.
XX
KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpr;
   vif; vpr; tat; rev; nef; vaccine.
XX
OS Human immunodeficiency virus 1.
XX
FN WO200026416-A1.
XX
PD 11-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US024837.
XX
PR 02-NOV-1998; 98US-00184418.
XX
PA (UABR-) UAB RES FOUND.
XX
PI Hahn BH, Shaw GM, Gao F;
XX
DR WPI; 2000-365651/31.
XX
PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
   type 1 useful for detecting and treating AIDS comprises a specific
   nucleotide sequence.
```

```
XX
PS Claim 41; Fig 14; 131bp; English.
XX
CC The present in invention provides the protein and coding sequences for a
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC isolates. The sequences shown include the near full-length coding
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC rev and nef proteins. These can be used to detect the presence of HIV-1
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC These antibodies can be used in vaccines to prevent and treat HIV
   infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 498 AA;
Alignment Scores:
Pred. No.: 7,64e-129 Length: 498
Score: 2404.00 Matches: 448
Percent Similarity: 94.22% Conservative: 25
Best Local Similarity: 89.24% Mismatches: 25
Query Match: 84.65% Indels: 4
DB: 3 Gaps: 1

US-09-475-704a-4 (1-1509) x AAB69276 (1-498)
QY 1 ATGGGCGCGCCGCGCGAGCATCTGCGCGCGCGAGAGTGGACAGTGGAGATCCGCG 60
   |||||
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyIleValSerLeuPheGlnIleArg 20
QY 61 CTGCGCGCGCGCGCGAGAGAGCATATCATCTGAGAGCATCTGTGTGGCGCCAGCGCGAG 120
   |||||
Db 21 LeuArgProGlyGlyIleValSerIleValSerMetLeuValSerIleValSerIleVal 40
QY 121 CTGAGAGGCTTGCCCTGAACCCCGCGCTGTGAGAGACCGCGAGGCGCTGCAAGCATC 180
   |||||
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuGluIleThrSerGlyGlyCysValSerGlnIle 60
QY 181 ATGAAGAGCTGACCGCGCGCGCTGAGAGCCGCGAGAGAGTGGAGCGCTGTACAC 240
   |||||
Db 61 IleGlnIleLeuGlnProAlaLeuGlnIleThrGlyIleValSerLeuIleValSerLeuIle 80
QY 241 ACCGTGGCGCACCTTACTGCTGCTGAGCGCGCGAGTGTGCGCGAGAGAGCGCC 300
   |||||
Db 81 ThrValAlaThrLeuIleCysValHisGlnIleValSerIleValSerIleValSerIleVal 100
QY 301 CTGACAAAGATGAGAGAGAGAGCAAAATGCTCCAGAGAGAGAGAGAGAGAGAGAG 360
   |||||
Db 101 LeuSerIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
QY 361 GCCAGCGCGAGGTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
   |||||
Db 121 ThrGlyGlyValSerGlnAsnIleProIleValGlnAsnLeuGlnIleValSerIleVal 140
QY 421 CACGAGCGCATAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
   |||||
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaThrValIleValIleGlnIleValSerGly 160
QY 481 TTCAAGCGCGAGGTATCCCAATGTTCAACCGCGCTGAGAGAGAGAGAGAGAGAGAGAG 540
   |||||
Db 161 PheAsnProGlyValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnAsn 180
QY 541 CTGAACAGAGATTGAACACCTGGGCGGCGCCAGCGCGCGAGAGAGAGAGAGAGAGAGAG 600
   |||||
Db 181 LeuAsnThrMetLeuAsnThrValGlyIleValIleValIleValIleValIleValIleVal 200
QY 601 ACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
   |||||
Db 201 ThrIleAsnGlnGlnAlaIleGlnIleThrPheArgIleValSerIleValSerIleValSer 220
QY 661 GCCCGCGCGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
   |||||
Db 221 AlaProGlyGlnIleArgGlnIleProArgIleValSerIleValIleValIleValIleVal 240
QY 721 CAGGAGAGAGATGCGCTGATATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
```



```

Db      241 GINGLIGLIIIEAIIATPMEETHRAASHANPROPILEPROVALGIIYAPILIEYRLYS 260
QY      781 CSGETGATCATCTGGAGCTGAACAAGATCTGCGGATGTACAGCCCGTGAGCATCTG 840
Db      261 AGTTRPIIEIIIELEUGIIYLEUENLYBIIEVALARGMETYSERPROVALSERIIIELEU 280
QY      841 GACATCCGCGAGGGGCCCCCAAGAGCCCTTCCGCGATCTAGGTGACCGCTTTCTTCAAGACC 900
Db      281 AAPILIELEGLINGIYPROLYSGIUPROPIEAARGSPYRVALAASPARGPHELYSTR 300
QY      901 CTGCGCGCCGAGCAGGCGCAGCAGCGAGGAGTGAAGATGTGATGACCGAGCCCGCTGGG 960
Db      301 LEUARGIAGIUGIINIAIATHRGINAPVALYS***TIPMETHTHASPHTXNEUVAL 320
QY      961 CAGAACCGCAACCCCGCATGTCAGAACCATCTGCGGCTCTCGGCCCCGAGCCACCTG 1020
Db      321 GINAEALIAASHPROBAPCYSLYSTHRIIELEUARGIALEUGIYPROGIIYAIATHRIE 340
QY      1021 GAGGAGATGATACCGCCCTGCGAGGCGCTGGGCGGCGCCGCGCAGAGCCCGCTGCTG 1080
Db      341 GIUGIUMETHTHALCYSGINGIYVALIGIYIYHISLYSIIAARGVALLEU 360
QY      1081 GCCGAGCGATGAGCCAGGCGCAAGCGTGAACATCATGATGACAGAAAGCAACTTCAAG 1140
Db      361 AIAGIUALIAMESECDINVALASHNANTHRAIILEMETGILNYSERASHNPHLYS 380
QY      1141 GGGCCCCGGCGCAAGCTCAAGTCTTCAACTGCGGCAAGAGGCGCCACATCGCCAGAAC 1200
Db      381 GLYPOLYIARGIIIEVALIYSCYSPHEANCYSGIYARGIUGIYHISIIIEALARGAEN 400
QY      1201 TGGCGGCGCCCGCGCAAGAGGCGCTGTGAAGTGGCGGAAGAGGCGCAAGATGAG 1260
Db      401 CYSAARGIAPROGIIYLYSGLYCYSTRPLYSYSGIYLYSGIUGIYHISGLIMETLYS 420
QY      1261 GACTGACCGAGCGCGCGCAACTCTCTGGCAAGATCTGGCGCCAGCAAGAGGCGCGC 1320
Db      421 AASPCTHRIIARGIINIAIASHNPHLEUGIYLSIIETRPROSECDINLYSGLIYARG 440
QY      1321 CCCGCGCACTTCTGTGCAAGACCGAGCGAGCGCGCGCCCGCCAGCGTGGCCAGCCGCC 1380
Db      441 PROGLIYASHNPHLEUGIINASHNARGPROGLIUPROTHRIIAPROTHRIIAPROTHRIIAPRO 460
QY      1381 CCCGCGAGAGCTTCCGCTTGTGAGAGACCAACCCCGCGCCCGCCAGAGAGGCCCAAGAC 1440
Db      461 PROIAGIUSERPHEARGPHEGIUGIUNHTRHTRPROIARGLIUGIINLYSASP 480
QY      1441 CGCGAGCCCTTACCGGAGCGCCCTGACCGGCTTGCAGCGCTTGTGGAGCGGCGCCCTG 1500
Db      481 -----LYSGIUPROLEUTHRIIALSEULYSERLEUPHEGLIYSEIASPPOLEU 496
QY      1501 AGCCGAG 1506
Db      497 SERGIN 498

RESULT 14
AAB69275
ID AAB69275 standard; protein; 494 AA.
XX
AC AAB69275;
XX
DT 12-SEP-2003 (revised)
DT 20-APR-2001 (first entry)
XX
DE HIV-1 non-subtype B clone 962M651-8 gag protein.
XX
KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW vif; vpr; tat; rev; nef; vaccine.
XX
OS Human immunodeficiency virus 1.
XX
PN WO200026416-A1.
XX

```

```

PD      11-MAY-2000.
XX
XX      25-OCT-1999; 99WO-US024837.
XX
XX      02-NOV-1998; 98US-00184418.
XX
XX      (UABR-) UAB RES FOUND.
XX
XX      Hahn BH, Shaw GM, Gao F;
XX
XX      WPI: 2000-365651/31.
XX
XX      Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT      type 1 useful for detecting and treating AIDS comprises a specific
PT      nucleotide sequence.
XX
XX      Claim 41; Fig 14; 131pp; English.
XX
XX      The present invention provides the protein and coding sequences for a
CC      number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC      isolates. The sequences shown include the near full-length coding
CC      sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC      rev and nef proteins. These can be used to detect the presence of HIV-1
CC      in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC      These antibodies can be used in vaccines to prevent and treat HIV
CC      infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
XX      SQ Sequence 494 AA;
XX
XX      Alignment Scores:
XX      Pred. No.: 9,92e-129 Length: 494
XX      Score: 2402.00 Matches: 455
XX      Percent Similarity: 93.85% Conservative: 18
XX      Best Local Similarity: 90.28% Mismatches: 19
XX      Query Match: 84.58% Indels: 12
XX      DB: 3 Gaps: 3
XX
US-09-475-704a-4 (1-1509) x AAB69275 (1-494)
QY      1 ATGGGCGCGCCGCGCAGCATCTCGCGCGAGAGCTGCAAGATGGAGAGATCCGC 60
Db      1 METGIYAIARGIASERIIELEUARGIYGIYLYSLEUAPLYSTRPGIILYIIAARG 20
QY      61 CTGGCGCCCGCGCGCAGAGCACTACTACATGTGAAGCACTGTGTGGCGCCAGCGGAG 120
Db      21 LEUARGPROGIYGIYLYSARGTYMETIIEYSHISEUVAITTPALASEIARGLI 40
QY      121 CTGAGGGGCTTGGCCCTGAACCCCGGCGCTGTGAGACCGCGAGGCGCTGCAAGCATC 180
Db      41 LEUINARGPHEALALEUASHNPROGIYLEUENGIUNHTRSERGIUGIYCYSLYSGIINII 60
QY      181 ATGAAGCAGTGCAGCCCGCGCTTGACAGCCGCGACCGAGAGAGTGCAGACTGTACAC 240
Db      61 METLYSGIINLEUGIINPROIALSEUGIINTHRCIYTHRIUGIINLEUARGSERIETRYASN 80
QY      241 ACCGTGGCAACCCCTGACTGTGTGACAGCGCGCATGTGAGTCCGCGCAGCAAGAGAGGCC 300
Db      81 THRVALIATHRIEUTRYCYSVAIHISGLIUGIYVALIGIUNALAGAPTRHLYSGIUALA 100
QY      301 CTGAGCAGATTCAGAGAGGAGCAGAACAGTCCAGCAGAGAGAACCCAGAGGCGCAAGAG 360
Db      101 LEUAPARGIIEGIUGIUGIUGIINASHNLYSIIIEGLINLYSIIIEGLINLYSTHRI 120
QY      361 -----GCCGACGGCAAGGTGAGCCAGAACTACCCCATGTGTGAGAACTTGACGGCCAG 414
Db      121 GINAIALIASPGLIYLSVALISERGINASHNTRYPROIIEVALGINASHNLEUGINGI 140
QY      415 ATGGTGACCAAGGCGCAGCGCCCGCAGACCTGAAGGCTGGGTGAAGGTGATGAGAGAG 474
Db      141 METVALHISGLIINLYSLEUSERPROARGTHREUASHNIAATRPVALIYVALIIEGLIG 160
QY      475 AAGGCTTTCAGCCCGGAGGTATCCCATGTTCACCGCCCTGAGCGAGGCGCCAGCCGCC 534

```



Db 161 LysAlaPheSerProGluValIlePrometPheThrAlaLeuSerGluGluAlaThrPro 180  
QY 535 CAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCACAGGCCCGCCATGCATGCTG 594  
Db 181 GlnAspLeuAsnThrMetLeuAsnThrValGlyGlyIleGlnAlaAlaMetGlnMetLeu 200  
QY 595 AAGGACACATCAAGAGGAGGCGCGGAGTGGGAGCCGCTGACCCCGGAGCGGCGG 654  
Db 201 LysAspThrIleAsnGluGluAlaAlaGluTrpAspArgLeuHisProValHisAlaGly 220  
QY 655 CCCGTGGCCCGCCGAGATGCGGACCCCGCGGAGCGACATCGCCGCGCCACGAG 714  
Db 221 ProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSer 240  
QY 715 ACCCTGCAGAGACGATCGCTTGGATGACCAACGACCCCGCTGGCCGCGGACATC 774  
Db 241 ThrLeuGlnGluGlnIleAlaThrMetThrSerAsnProIleProValGlyAspIle 260  
QY 775 TACAAGCGGTGATCATCTGCGGCTGGAACAAGATGTCGGATGTCACGCGCCGTGAG 834  
Db 261 TyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSer 280  
QY 835 ATCTGTGAATCCGCGGAGGCGCCCAAGAGCCCTTCGCGACATACGTGACCGCTTCT 894  
Db 281 IleLeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhe 300  
QY 895 AAGACCTTGGCGCGCCGAGCAGCGCCACCCAGAGCGTGAAGAATGTGATGACCGAGACTG 954  
Db 301 LysThrLeuArgGlnGluGlnAlaThrGlnGluValLysAsnThrMetThrAspThrLeu 320  
QY 955 CTGGTGCAGAACCCCAACCCCGCATGCACACATCTGCGGCGCTTCCGCGCCCGCGCC 1014  
Db 321 LeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProGlyAla 340  
QY 1015 ACCCTGAGAGATGATGACCGCTGCGAGCGGTGGCGCGCCCGCCGACAGAGCCCGC 1074  
Db 341 ThrLeuGlnGluMetThrAlaCysGlnGlyValGlyLysProSerHisLysAlaArg 360  
QY 1075 GTGCTGGCGCGGCGATGAGCCAGGCGCCACAGCGTGAACATCATGATGACAGAGCAAC 1134  
Db 361 ValLeuAlaGlnAlaMetSerGlnThrAsnSerValAsnIleLeuMetGlnLysSerAsn 380  
QY 1135 TTCAGAGGCGCCCGGCGCACTGCATGCTTCAATGCGCGAGAGAGGCGCCATCGCC 1194  
Db 381 PheLysGlyAsnLysArgMetValLysCysPheAsnCysGlyLysGluGlyHisIleVal 400  
QY 1195 AAGAACTGCGCGCGCCCGCGCAAGAGGCGTGTGGAAGTGGCGAGAGGCGCCACGAG 1254  
Db 401 ArgAsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGluGlyHisGln 420  
QY 1255 ATGAAGAGATGCAACCGAGCGCCAGGCGCACTTCTGCGGCAAGATCTGGCCAGCCAGAG 1314  
Db 421 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 440  
QY 1315 GGGCGCGCGCGCACTTCTGCGCAAGACCGAGCGAGCCCGCGCCCGCCAGCGCCAGC 1374  
Db 441 GlyArgProGlyAsnAsnLeuGlnAsnArgProGlu-----ProThr 454  
QY 1375 GCGCGCGCGCGCGAGCTTCCGCTTCAGAGAGACCAACCCCGCCCAAGCAGAGAGCC 1434  
Db 455 AlaProProAlaIleSerPheArgPheGluGluThrThrProAlaProLysGlnGlnSer 474  
QY 1435 AAGAGCGGAGAGCCCTTACCGGAGCCCTGACCGCCCTGCGAGCGCTGTTCCGAGCGAG 1494  
Db 475 LysAsp-----ArgGlnAlaLeuThrSerLeuLysSerLeuPheGlySerArg 490  
QY 1495 CCCCTAGCGCAG 1506  
Db 491 ProLeuSerGln 494

RESULT 15  
ADCT2875  
ID ADCT2875 standard; protein; 508 AA.

XX AC ADCT2875;  
XX DT 18-DEC-2003 (first entry)  
XX DE HIV-1 gag protein containing an NCP sequence, SEQ ID NO 105.  
XX KW complex; HIV nucleocapsid protein 7; NCP7; HIV-ps1-site; anti-HIV;  
XX KW human immunodeficiency virus; HIV.  
XX OS Human immunodeficiency virus 1.  
XX PN WO2003060998-A2.  
XX PD 24-JUL-2003.  
XX PF 10-JAN-2003; 2003WO-US000801.  
XX PR 11-JAN-2002; 2002US-0347369P.  
XX PA (AChI-) ACHILLION PHARM INC.  
XX PI Bechter D, Hou X, Marlor CW, Rice WG, Yang W;  
XX DR WPI; 2003-646042/61.  
XX PT Method of determining whether a compound inhibits formation of complex  
XX PT between HIV nucleocapsid protein 7 polypeptide and HIV ps1-site  
XX PT oligonucleotide by comparing amount of complex formed in presence/absence  
XX PT of compound.  
XX PS Claim 2; SEQ ID NO 105; 105pp; English.  
XX CC The invention relates to a novel method of determining whether a compound  
XX CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)  
XX CC polypeptide and an HIV-ps1-site oligonucleotide. The method involves  
XX CC adding the HIV-ps1-site oligo to a mixture of the NCP7 polypeptide and  
XX CC the novel compound and comparing amount of complex formed, with that  
XX CC formed in the absence of the compound, where a decrease in the amount of  
XX CC complex formed in presence of the compound indicates that the compound  
XX CC inhibits complex formation. The method of the invention involves anti-HIV  
XX CC activity. The compounds identified are useful for treating a subject  
XX CC infected with human immunodeficiency virus (HIV) by administering the  
XX CC compound to the subject. This sequence represents an HIV-1 gag protein  
XX CC which contains an NCP sequence of the invention.  
XX SQ Sequence 508 AA;  
XX Alignment Scores:  
XX Pred. No.: 7,11e-128  
XX Score: 2387.00 Length: 508  
XX Percent Similarity: 92.19% Matches: 454  
XX Best Local Similarity: 88.67% Mismatches: 18  
XX Query Match: 84.05% Indels: 26  
XX DB: 7 Gaps: 14  
XX US-09-475-704A-4 (1-1509) x ADCT2875 (1-508)  
QY 1 ATGGAGCGCCGCGCGAGCATCTGCGCGGCGAGAACTGAGACAGTGGAGAAATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyThrLysLeuAspAlaTrpIleLysIleArg 20  
QY 61 CTGGCGCGCGCGCGCAAGAGCATACATGCTGAAGACACTGTGTGTGGCCAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysHisTyrMetLeuLysIleValTrpAlaSerArgGln 40  
QY 121 CTGGAGGCGTTCGCGCTGAACCCCGCGCTGCGAGACCGCGAGGCGTGCAGAGATC 180  
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLysGluThrThrSerGlnGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGACGCGCCCTGACAGACCGGACCAAGAGAGCTGCGAGCTGTACAAC 240  
Db 61 MetLysGlnLeuHisProAlaLeuGlnThrGlyThrGlnGluLeuLysSerLeuTyrAsn 80

QY 241 ACCGTGGCCACCGCTGACTGCGTGCACCGCGGCGATGAGAGTCCGGGACACCAAGAGAGCC 300  
 DB ThrValAlaThrLeuValCysValHisGluAsnIleLeuValArgAspThrIysGluAla 100  
 QY 301 CTGCAACAAGTACGAGGAGGAGGAGCAAC-----AAGTCCACAGAGAACCCGACGAGCC 354  
 DB LeuAspIysIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120  
 QY 355 AAGAGGCGCGAGCGGCAAGGTGAGCCGAACCTAACCCCATGTGCGAGAACTCGAGAGGCGAG 414  
 DB LysAlaAlaAspGluIleValSerGlnAsnIleValGlnAsnLeuGlnGluGln 140  
 QY 415 ATGGTGACACGAGGCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474  
 DB MetValHisGlnAsnLeuSerProAlaGlnLeuAsnAlaIleProValIleGluGlu 160  
 QY 475 AAGGCTTTCAGCGCGCGAGGTGATCCCATGTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 534  
 DB LysValAspSerProGluValIleProMetPheThrAlaLeuSerGluIleAlaThrPro 180  
 QY 535 CAGAACCTGGAACACGATGTTTAAACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 594  
 DB GlnAspLeuSerThrMetLeuAsnThrValGluGluGlnAlaAlaMetGlnMetLeu 200  
 QY 595 AAGGACACCATGACAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 654  
 DB LysAspThrIleAsnGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 220  
 QY 655 CCCGTGGCGCGCGCGCGAGTGCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 714  
 DB PrometAlaProGluGluLeuArgIleProAlaGlySerAspIleAlaGluThrThrSer 240  
 QY 715 ACCCTGCAGAGACAGATCGCTGATGACAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 774  
 DB ThrLeuArgGluGlnIleAlaIleThrMetThrSerAsnProIleProValGluAspIle 260  
 QY 775 TACAGCGGCTGATGATCTGTGGCGCTGAACAGATGCTGCGATGACAGCGCGCGCGCGCGCG 834  
 DB TyrLysAspGlyPheIleLeuGluLeuAsnIleValArgMetCysSerProValSer 280  
 QY 835 ATCTGGAACATCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 894  
 DB IleLeuAspIleArgGlnGluProLysGluProPheArgAspTyrValAspArgPhePhe 300  
 QY 895 AAGACCTTGCAGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954  
 DB LysAlaLeuArgAlaGluGlnIleAlaThrGlnAspValLysAsnIlePheThrGluThrLeu 320  
 QY 955 CTGTGCGAGAACCGCAACCGCGAGTGCAGACCATCTGCGCGCGCTGCGCGCGCGCGCGCGCG 1014  
 DB LeuValGlnAsnIleAsnProAspCysLysIleLeuLysAlaLeuGluIleGluAla 340  
 QY 1015 ACCCTGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074  
 DB ThrLeuGluGluMetMetThrAlaCysGlnGluValGluGlyProSerHisLysAlaArg 360  
 QY 1075 GTGCTGGCGCGAGCGCATGAGCGCGCAACAGCGTGAACATGATGATGCGAAGAGCAAC 1134  
 DB ValLeuAlaGluAlaMetSerGlnIleAsnAsnThrAsnIleMetMetGlnArgSerAsn 380  
 QY 1135 TTCAAGGGGCGCGCGCGCAACGTTCAAGTGTCTCAACTGCGGAGAGGAGGCGCGCAATCGCGC 1194  
 DB PheLysSerSerLysArgIleValLysCysSerAsnCysGlyLysGluIleHisIleAla 400  
 QY 1195 AAGAACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGGCGGAGAGGCGCGCGCGCGCGCG 1254  
 DB ArgAsnCysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGluGluGln 420  
 QY 1255 ATGAAGACTGACCGAGGCGCGCGCAACTTCTGGGCGAAGATCTGGCGCGCGCGCGCGCGCGCG 1314  
 DB MetLysAspCysThrGluArgGlnIleAsnPheLeuGluLysIleTyrProSerHisLys 440

QY 1315 GAGCGCGCGCGCACTTCTGCGAGAACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1368  
 DB GlnArgProGluAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGluSer 460  
 QY 1369 -----CCACCGCGCGCGCGCGCGAGAGCTTCCGCTTCGAGGAGAC 1410  
 DB PheArgAsnArgProGluProThrAlaProProAlaGluSerPheArgPheGluGluThr 480  
 QY 1411 ACCCGCGCGCGCGCGAGAGCGCGCAAGACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1470  
 DB ThrProThrProLysGlnGluProLysAsp-----ArgAspProLeuThrSer 496  
 QY 1471 CTGCGAGCGTGTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1506  
 DB LeuLysSerLeuPheGlySerAspProSerSerGln 508

Search completed: March 11, 2005, 15:22:46  
 Job time : 241.806 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 14:56:45 ; Search time 259.58 Seconds  
(without alignments)  
5953.668 Million cell updates/sec

Title: US-09-475-704A-4

Perfect score: 2840  
Sequence: 1 atgggcgcgcgcgcagcatc.....gcgcgcctctgagccagtaa 1509

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
Model=frame+ n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool.p/US09475704/runcat.10032005.140221.14821/app\_query.fasta\_1.3342  
-DB=uniprot.03 -QFMT=fastaan -SUFIX=n2p.rup -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human0.cdi -LIST=45  
-DOCLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704\_QCGN.1.1.61a@runat.10032005.140221.14821 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: uniprot.03:\*  
2: uniprot.trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2661	93.7	502	2	Q9WF78
2	2464	86.8	492	2	Q901X9
3	2458	86.5	500	2	Q90056
4	2457	86.5	492	2	Q8AFM4
5	2445.5	86.1	495	2	Q6XKD2
6	2440	85.9	492	2	Q8UT7A
7	2439.5	85.9	499	2	Q8USV8
8	2439	85.9	508	2	Q8USY5
9	2438.5	85.9	497	2	Q6S7X9
10	2437	85.8	492	2	Q90ME1
11	2435.5	85.8	491	2	Q90N98
12	2433	85.7	492	2	Q901X0
13	2430.5	85.6	491	2	Q7SLX7
14	2430	85.6	492	2	Q90127
15	2430	85.6	492	2	Q90ME2
16	2429	85.5	492	2	Q6S859

17	2429	85.5	500	2	Q8UT80
18	2428.5	85.5	491	2	Q7SLX1
19	2427.5	85.5	491	2	Q90090
20	2427	85.5	492	2	Q8UTK6
21	2425.5	85.4	491	2	Q8AFM1
22	2425	85.4	488	2	Q7SLY2
23	2424	85.4	500	2	Q90ME7
24	2422.5	85.3	491	2	Q90081
25	2421.5	85.3	491	2	Q90064
26	2421.5	85.3	491	2	Q7SLY0
27	2421.5	85.3	491	2	Q7SLY6
28	2421.5	85.3	491	2	Q9Q3C6
29	2419	85.2	492	2	Q90MF5
30	2419	85.2	492	2	Q9WF93
31	2417.5	85.1	501	2	Q8UTD4
32	2417	85.1	492	2	Q9WF90
33	2417	85.1	500	2	Q8UTN3
34	2416.5	85.1	499	2	Q994H3
35	2415	85.0	492	2	Q6S7X0
36	2415	85.0	492	2	Q6S883
37	2414.5	85.0	495	2	Q8USZ4
38	2414.5	85.0	499	2	Q6X6H4
39	2414	85.0	492	2	Q90MF4
40	2414	85.0	492	2	Q9WF41
41	2413.5	85.0	491	2	Q90072
42	2413.5	85.0	499	2	Q8UTN9
43	2413	85.0	492	2	Q9WF63
44	2412.5	84.9	503	2	Q7SQ46
45	2412	84.9	492	2	Q9WF87

## ALIGNMENTS

RESULT 1  
Q9WF78 ID Q9WF78 PRELIMINARY: PRT: 502 AA.  
AC Q9WF78  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Gag polyprotein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Nung'u T.P., Rahman M., Makheia M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana".  
RT J. Virol. 73:4427-4432(1999).  
RL EMBL, AF110967; AAD17071.1; --  
DR HSP; P24736; INCP.  
DR GO; GO:0019012; C:Virusion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR00721; Gag\_p24.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR001071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR InterPro; IPR01878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; I.  
DR Pfam; PF00607; Gag\_p24; I.  
DR Pfam; PF00098; Zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.





RT C-infected seroconverters in India, with evidence of intersubtype recombination";  
 RL J. Virol. 73:152-160(1999).  
 DR EMBL; AF067154; AAD12071.1; -.  
 DR HSP; Q70622; IHN.  
 DR GO; GO:0019012; C:vixion; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR010999; Retrovir\_matrix.  
 DR InterPro; IPR000071; Retrovir\_p17.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR008919; Retrov\_capsid\_N.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF00098; Zf\_CCHC; 2.  
 DR PRINTS; PR00939; G2HCZNFINGER.  
 DR PRINTS; PR00234; HIVMATRIX.  
 DR SMART; SM00343; ZNF\_C2HC; 2.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KM AIDS; Core protein; Metal-binding; Polypeptide; Zinc; Zinc-finger.  
 SQ SEQUENCE 500 AA; 55878 MW; 25AA23024004E325 CRC64;

## Alignment Scores:

Pred. No.: 3,46e-98 Length: 500  
 Score: 2458.00 Matches: 465  
 Percent Similarity: 95.45% Conservative: 17  
 Best Local Similarity: 92.08% Mismatches: 15  
 Query Match: 86.55% Indels: 8  
 DB: 2 Gaps: 3

US-09-475-704A-4 (1-1509) x 090056 (1-500)

QY 1 ATGGGCGCCGCGCGCAAGATCTTGGCGCGCGCAAGCTGGACATGGAGAAATCCGC 60  
 Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspLysTrpGluArgIleArg 20  
 QY 61 CTGCGCGCCGCGCGCAAGATCATGTGTGAGACCTGTGTGGCGCGCGCGAG 120  
 Db 21 LeuArgProGlyGlyLeuLysHisIleMetLeuLysHisLeuValTrpAlaSerArgGlu 40  
 QY 121 CTGAGAGGCTTGTGCTGAACTCCGCGCTGTGTGAGACCGCGAGGCTGTCAAGCATC 180  
 Db 41 LeuAspArgPheAlaLeuAsnProGlyLeuLeuGluAlaIleGlyCysIleGlnIle 60  
 QY 181 ATGAGCAGCTGTGAGCGCGCGCTGTGAGACCGCGAGAGCTGTGCGCTGTACAC 240  
 Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuHisAsn 80  
 QY 241 ACCGTGCGCGCTGTGCTGTGAGACCGCGCATGTGAGTCCGCGCACCAAGAGAGCC 300  
 Db 81 ThrValAlaThrLeuIleCysValHisGluGlyIleGluValArgAspThrLysGluAla 100  
 QY 301 CTGGAACAAGTGTGAGAGAGAGCAAGCAAGTCCGCGAGCAAGAGAGAGAGAGAG 360  
 Db 101 LeuAspLysIleGluGlnGluGlnAsnLysSerGlnGlnLysTrpGlnGlnAlaLysGlu 120  
 QY 361 GCCGAGCGCAAGGTGAGCGAGCAAGCAAGCTGTGTGAGAGCTGTGAGAGAGAGAGAG 420  
 Db 121 AlaAspGlyLysValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlyMetVal 140  
 QY 421 CACGAGCGCATCAAGCCCGCGAGCTGTGAGAGCTGTGAGAGAGAGAGAGAGAGAG 480  
 Db 141 HisGlnProIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLysAla 160  
 QY 481 TTGAGCCCGAGGATGCCATGTTCACCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAG 540  
 Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
 QY 541 CTGAACAAGATGTGAACACCGTGGCGCGAGCGAGCGCGCATGTGAGAGCTGTGAAGAC 600

Db 181 LeuAsnThrMetLeuAsnThrValGlyGlnIleAlaIleMetGlnMetLeuLysAsp 200  
 QY 601 ACCATCAACGAGAGAGCGCGAGTGGAGCGCTGTGACCCCGTGCAGCGCGCGCGCTG 660  
 Db 201 ThrIleAsnGlnGluAlaIleGluTrpAspArgLeuHisProValHisAlaGlyProIle 220  
 QY 661 GCCCGCGCGAGATCGCGAGAGCGCGCGCGCGAGCATCGCGCGCGCGAGAGAGAGAG 720  
 Db 221 AlaProGlyGlnIleArgLysProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
 QY 721 CAGAGCAGATGTGCTGTGATGACCAAGACCGCGCGCGCGCGCGAGATCTGTACAG 780  
 Db 241 GlnGlnGlnIleAlaTrpMetThrGlyAsnProProValProValGlyAspIleTrpLys 260  
 QY 781 CGGTGATCATCTGTGGCTGTGAACAAGTGTGTGGATGTACAGCCCGGTGAGATCTGTG 840  
 Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetCysSerProValSerIleLeu 280  
 QY 841 GACATCGCGCGAGCGCGCGCGAGAGCGCTTCCGCGCATTCGTGAGACCGCTTTCAGAGC 900  
 Db 281 AspIleLysGlnGlyProLysGlnProPheArgAspTrpValAspArgPheArgThr 300  
 QY 901 CTGCGCGCGCGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 Db 301 LeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320  
 QY 961 CAGAACGCGCAACCGCGAGCTGTGAGACCATCTGTGCGCTGTGCGCGCGCGAGAGAG 1020  
 Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
 QY 1021 GAGAGATGATGACCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 Db 341 GlnGlnMetMetThrAlaCysGlnGlyValGlyProGlyLysLysValAlaArgValLeu 360  
 QY 1081 GCGAGCGAGATGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 Db 361 AlaGlnAlaMetSerGlnAlaAsnSer--AsnIleuMetGlnArgSerAsnPheLys 379  
 QY 1141 GCGCGCGCGCGCAACGTCATGTCTTCACTGTGCGCGAGAGAGAGAGAGAGAGAGAG 1200  
 Db 380 GlySerLysArgThrValLysCysPheAsnLysGlyLysGlnGlyHisIleAlaArgAsn 399  
 QY 1201 TGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 Db 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisIleMetLys 419  
 QY 1261 GACTGCACGAGAGCGCGCAACTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 Db 420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439  
 QY 1321 CCGCGCAACTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371  
 Db 440 ProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaArgProGluPro 459  
 QY 1372 ACCGCGCGCGCGCGAGAGCTTCCGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431  
 Db 460 ThrAlaProProAlaGlnSerPheArgPheGlnGlnThrThrProAlaLeuLysGlnGlu 479  
 QY 1432 CCGAGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1491  
 Db 480 ProLysAsp-----ArgGluProLeuThrSerLeuLysSerLeuPheGlySer 495  
 QY 1492 GCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506  
 Db 496 AspProLeuSerGln 500

## RESULT 4

Q8AFW4 PRELIMINARY; PRT; 492 AA.  
 ID Q8AFW4;  
 AC Q8AFW4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Gag protein.  
 GN Name-gag;  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCB1\_TaxID=11676;  
 OX (1)  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2525961; PubMed=12639249; DOI=10.1089/08992203762686649;  
 RA Williamson C., Morris L., Naughtan M.F., Ping L.-H., Dryga S.A.,  
 RA Thomas R., Reap E.A., Cilliers T., van Harmelen J., Pascual A.,  
 RA Ramjee G., Gray G., Johnston R., Karim S.A., Swanstrom R.;  
 RT "Characterization and selection of HIV-1 subtype C isolates for use in  
 RT vaccine development."  
 RL AIDS Res. Hum. Retroviruses 19:133-144(2003).  
 DR EMBL: AF544010; ANO16234.1; -.  
 DR HSP: Q70622; 1HVN.  
 DR GO: GO:0019012; C:vitrion; IEA.  
 DR GO: GO:0003676; P:nucleic acid binding; IEA.  
 DR GO: GO:0005198; P:structural molecule activity; IEA.  
 DR GO: GO:0016032; P:viral life cycle; IEA.  
 DR InterPro: IPR000721; Gag\_p24.  
 DR InterPro: IPR010999; Retrovir\_matrix.  
 DR InterPro: IPR000071; Retrovir\_p17.  
 DR InterPro: IPR008916; Retrov\_capsid\_C.  
 DR InterPro: IPR008919; Retrov\_capsid\_N.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF00540; Gag\_p17; I.  
 DR Pfam: PF00607; Gag\_p24; I.  
 DR Pfam: PF00098; zF-CCHC; 2.  
 DR SMART: SM00343; ZNF\_C2HC; 2.  
 DR PROSITE: PS0158; ZF\_CCHC; 2.  
 KW AIDS: Core protein; Polypeptide.  
 SQ SEQUENCE 492 AA; 55221 MM; B1922020A1AE861B CRC64;

## Alignment Scores:

Pred. No.: 3.83e-98 Length: 492  
 Score: 2457.00 Matches: 465  
 Percent Similarity: 95.22% Conservative: 13  
 Best Local Similarity: 92.63% Mismatches: 14  
 Query Match: 86.51% Indels: 10  
 DB: 2 Gaps: 2

US-09-475-704a-4 (1-1509) x Q8AFW4 (1-492)

QY 1 ATGGGGGCGCGCGCGCATCTCTGGCGCGGAGAGACTGTGACAACTGGAGAAATCCGC 60  
 DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyIuLeuAspLysTrpGluysIleArg 20  
 QY 61 CTGGCGCGCGGCGGAGAGAGACTGATGCTGAGACGACTGGGTGGGCGAGCGCGAG 120  
 DB 21 LeuArgProGlyGlyIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 40  
 QY 121 CTGGAGGGCTTGGCTGAGACCCGCGCTGTGTGAGACCGCGAGGGCTGTGACAGATC 180  
 DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluIuysIuysIuysIuysIuys 60  
 QY 181 ATGAAACAGCTGAGCGCGCGCTGTGAGACCGCGAGAGAGTGTGGCAGCTGTACAC 240  
 DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluIuysIuysIuysIuysIuysIuys 80  
 QY 241 ACCGGGCGACCTGTACTGCGTGTGAGCGCGGCGATGAGGTCCGGGACCAAGAGGGCC 300  
 DB 81 ThrValAlaThrLeuLysCysValIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 100  
 QY 301 CTGGACCAAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 DB 101 LeuAspLysIleGluIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 120  
 QY 361 GCGGAGCGGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
 DB 121 AlaAspGlyLysValSerGlnAsnThrProIleValGlnAsnLeuGlnIuysIuysIuysIuysIuys 140  
 QY 421 CACGAGGCGATCAGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480

DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValIuysIuysIuysIuysIuysIuysIuys 160  
 QY 481 TTCAGCGCGGAGAGTGTATCCCATGTTTACCGCGCTGTGAGAGAGGCGGACCGCGGAGAG 540  
 DB 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluIuysIuysIuysIuysIuysIuysIuys 180  
 QY 541 CTGAACAGAGTGTGAACACCGTGGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 DB 181 LeuAsnThrMetLeuAsnThrValGlyIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 200  
 QY 601 ACCATCAACAGAGAGAGCGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
 DB 201 ThrIleAsnGlu\*\*\*AlaIleGluTrpPheArgLeuHisIleProValHisIleGlyProIle 220  
 QY 661 GCGCGCGCGGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 DB 221 AlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeu 240  
 QY 721 CAGGAGAGATGCGCTGATGACAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 DB 241 GlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTrpLys 260  
 QY 781 CGGTGATCATCTGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
 DB 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValAlaGlyMetLysSerProValSerIleLeu 280  
 QY 841 GACATCCGCGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 DB 281 AspIleArgGlnGlyProLysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 300  
 QY 901 CTGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 DB 301 LeuArgAlaGluGlnAlaThrGlnIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 320  
 QY 961 CAGAGAGCGCAACCGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuAlaGlnIuysIuysIuysIuysIuysIuysIuys 340  
 QY 1021 GAGAGAGATGATGACGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 DB 341 GlnGluMetMetThrAlaCysGlnIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 360  
 QY 1081 GCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 DB 361 AlaGlnAlaMetSerGlnThrAsnSerGlyAsnIleMetMetGlnArgSerAsnPheLys 380  
 QY 1141 GCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 DB 381 GlyProAlaArgGlyIleValLysCysPheAsnGlyLysIuysIuysIuysIuysIuysIuysIuys 400  
 QY 1201 TGGCGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 DB 401 CysArgAlaProAlaArgLysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 420  
 QY 1261 GACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 DB 421 AspCysTrpGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerThrIleLysGlyArg 440  
 QY 1321 CCGGCGCACTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
 DB 441 ProGlyAsnPheLeuGlnAsnArgProIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 460  
 QY 1381 CCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 DB 465 ProAlaGlnLysPheArgPheGlnIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuysIuys 474  
 QY 1441 GCGGAGCGCTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 DB 475 -----ArgGluProLeuThrIleSerLeuLysSerLeuPheGlySerAspProLeu 490  
 QY 1501 AGCCAG 1506  
 DB 1501 -----







QY	1201	TGCGCGGCCCGCCCGCAGAGGGCTGTGGAGTCGGCGAAGAGGCGCCACAGATGAAG	1260
Db	401	CyBaGALaPArPaArGlyeSGLYcYstrPlySvGSLYvGLInGLYhNIGLImeClyS	420
QY	1261	GACTGACACGAGCGCGCAGGCGCAACTTCCTGGGGCAGAGTATGGCCGACCAAGAGGCGCG	1320
Db	421	AapCyStrhGInaArgGLInAlaasnPhneugLYvLlEttrProSerhLaLyGLYArg	440
QY	1321	CCGGCAACTTCCTGCAGAACCGCAGCGACCCCGCGCCACCGTGGCCACGCGCCCG	1380
Db	441	ProGlyAasnPhLeuGLInSerArgProGlu-----ProThrAlaPro	454
QY	1381	CCGGCGGAGAGCTTCGCTGTGAGAGACACCCCGCCCGCCAGACGAGCCCAAGAC	1440
Db	455	ProAaGLInSerPhaArgPhneGLInuThrThrProAlaProLYsGLInGLuProLYaAsp	474
QY	1441	CGCAGACCCCTACCGGAGCGCCCTGACCGCCCTGCGCAGACTGTTCGGAGCGCGCCCTG	1500
Db	475	-----ArgGLuProLeuThrSerLeuLYsSerLeuPhneGLYSerAspProLeu	490
QY	1501	AGCCGAG 1506	
Db	491	Sergin 492	
RESULT 7			
Q8USV8	ID	Q8USV8	PRELIMINARY; PRT; 499 AA.
AC	08USV8;	01-MAR-2002 (TREMBLrel. 20, Created)	
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)		
DE	Gag protein.		
GN	Name=gag;		
OS	Human immunodeficiency virus 1.		
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
NCBI	taxid:11676;		
NCBI	taxid:11676;		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=21988475; PubMed=11991972;		
RX	DOI=10.1128/JVI.76.11.5435-5451.2002;		
RA	Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,		
RA	Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,		
RA	Foley B.T., GaoLekwe S., Rybak N., Gassetts S., Vannberg F.,		
RA	Marlink R., Lee T.-H., Essex M.;		
RT	"Human immunodeficiency virus type 1 subtype C molecular phylogeny:		
RT	consensus sequence for an AIDS vaccine design?";		
RL	J. Virol. 76:5435-5451(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,		
RA	Chang S.-Y., Peter T., Thior I., Rybak N., Gassetts S., Vannberg F.,		
RA	Marlink R., Lee T.-H., Essex M.;		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF431112; ALJ34891.1; -		
DR	HSSP; P24736; INCP.		
DR	GO; GO:0019012; C:viral; IEA.		
DR	GO; GO:0003676; F:nucleic acid binding; IEA.		
DR	GO; GO:0005198; F:structural molecule activity; IEA.		
DR	GO; GO:0016032; P:viral life cycle; IEA.		
DR	Pfam; PF00540; Gag_p17; 1.		
DR	Pfam; PF00607; Gag_p24; 1.		
DR	Pfam; PF00098; zf-CCHC; 2.		
DR	SMART; SM00343; znf_C2HC; 2.		
DR	PROSITE; PS0158; ZF_CCHC; 2.		
KW	AIDS; Core protein; Polyprotein.		
SO	SEQUENCE 499 AA; 55681 MW; 0BF837A38CA5779E CRC64;		

Query Match:	85.90%	Indels:	5
DB:	2	Gaps:	2
US-09-475-704A-4 (1-1509) x Q8USV8 (1-499)			
QY	1 ATGGGGGCGCCGGCCAGCATCTCGGGCGGCGAGAACCTGTGACAAAGTGGAGAGATCCGC	60	
Db	1 MetGlyAlaArgAlaSerIleLeuHdgLyGlyLysLeuAspLysTrpGluLysIleArg	20	
QY	61 CTGGCGCCCGCGCGCAAGACCTACATGCTGAACACCTGGTGTGGCCAGCCCGCAG	120	
Db	21 LeuArgProGlyGlyLysLysCysTrpMetIleLysHsIleIleLeuTrpLaserArgIle	40	
QY	121 CTGGAGGGGTTGGCCCTGAAACCCCGGCTGGCTGGAGAACCCCGAGGGCTGCAAGCATC	180	
Db	41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrAlaGluGlyCysThrGlnIle	60	
QY	181 ATGAAGCAGCTCAGCCCGCCCTGGCAGACCGGCAACCGAGACCTGGCAGCCTGTACAC	240	
Db	61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuPheAsn	80	
QY	241 ACCGTGGCCACCTCTGACTGCGTGCACCGCGCATCGAGTCCGCGACCAAGAGGCC	300	
Db	81 ThrValAlaThrLeuTrpCysValHisAlaGlyIleAspValAlaArgSerThrLysGluAla	100	
QY	301 CTGGAACAAGATGAGAGAGAGACAGAACTACCCCATCGTGCAGAACCTGACGGCCAGAGAG	360	
Db	101 LeuAspLysIleGluGluGluGluGlnAsnLysSerGlnGlnLysThrGlnGlnAlaGlnAla	120	
QY	361 GCCAGCGGCAAGTGAAGCCAGAACTACCCCATCGTGCAGAACCTGACGGCCAGATGTGTG	420	
Db	121 AlaAspLysLysValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlyGlnMetVal	140	
QY	421 CACCAAGGCATATAGCCCCCGGACCCCTGAAACGCTGGGGTGAAGTATGAGAGAAAGGCC	480	
Db	141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLysAla	160	
QY	481 TTCAGCCCCGAGTGATCCCATGTTTCAACCGGCTGACCGGCTGACGAGGCGCCACCGGAC	540	
Db	161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp	180	
QY	541 CTGAACACAGATGTTGAACACCGTGGGGCGGCCACAGGCGCCCATGACAGATGCTGAAGAC	600	
Db	181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp	200	
QY	601 ACCATCAACAGAGAGGCGCGGAGTGGGAGACCGGCTGACCGGCTGACGAGCGGCGCCGG	660	
Db	201 ThrIleAsnGluGlnAlaHisGluThrAspArgLeuHisProValHisAlaGlyProIle	220	
QY	661 GCCCGGCGCGATGCGCGACCCCGCGCGGACGACATGCGCGGCGGCCACGACCCCTG	720	
Db	221 AlaProGlyGlnMetArgGlnProArgIlySerAspIleAlaGlyThrThrSerThrLeu	240	
QY	721 CAGGAGCAGATCGGCTGTGATGACAGCAACCCCGGCGCCCTGGGGGAGACATCTACAG	780	
Db	241 GlnGluGlnIleAlaTrpMetThrAsnAsnProProValProValGlyAspIleLys	260	
QY	781 CGGTGATCATCTGGGCTGACAGAAAGATCGTGGGATGTACAGCCCGGTAGCATCTGT	840	
Db	261 ArgThrIleIleLeuGlyLeuAsnLysIleValAlaArgMetLysSerProValSerIleLeu	280	
QY	841 GACATCCGCGCCGCGGCCCAAGAGACCTTCGCGCACTAGTGAACCGCTTCTTCAAGACC	900	
Db	281 AspIleLysGlnGlyProLysGlnProPheArgAspTrpValAlaAspArgPheAspArgThr	300	
QY	901 CTGGCGCCCGGAGCAGGCCACCCAGAGAGGTGAAGATCTGAGACCGGACCTGTGGTGTG	960	
Db	301 LeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeuVal	320	
QY	961 CAGAAAGCCAAACCCGCACTGCAAGACCATCTGGCGGCTCTGGCGCCGCGGCAACCTGT	1020	
Db	321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThrLeu	340	





QY 721 CAGAGACAGATGCGCTGATGACACGACACCCCGCTGCGCCGACATCTACAG 780  
DB 241 GINGLUGLHILELITRPMETHTGlyAsnProProValProValGlyAspIleTyrLys 260  
QY 781 CCGTGATCATCTCGGCGCTGAAACAAGATGTCGATGACAGCCCGGACATCTCG 840  
DB 261 AGTTPRIleIleuLysLeuAsnLysIleValArgMetTyrSerProValSerIleu 280  
QY 841 GACATCGCGCAGGCGCCCAAGAGCCCTTCGCGACTGACGCGCTTCTTCAAGCC 900  
DB 281 AAPleArgGlnGlnIlyProLysGlnProPheArgAspTyrValAspArgPheLysThr 300  
QY 901 CTGCGCGCGCAGGCGCCCAAGAGCCCTTCGCGACTGACGCGCTTCTTCAAGCC 960  
DB 301 LeuArgLysGlnGlnIleValThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320  
QY 961 CAGAACGCGCAGCCCGCTGACAGACATCTCTGCGCTCTGCGCCCGCGACCTG 1020  
DB 321 GlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSerLeu 340  
QY 1021 GAGAGATGATGACCGCTGCGCAGGCGCTGCGCGCCCGCAGAGCCCGCTGCTG 1080  
DB 341 GlnGlnMetMetThrLysAlaCysGlnGlnIlyValIlyGlyProGlyIlyLysAlaArgValLeu 360  
QY 1081 GCGGAGCGCATGACGCGCCCAAGAGCCCTGACATCTGATGACAGAGAGCACTTCAAG 1140  
DB 361 AlaGlnAlaMetSerGlnAlaAsnSer---AsnIleMetMetGlnArgGlyAsnPheLys 379  
QY 1141 GAGCGCGCGCAGACGCTCAAGTGTCTCAACTGCGCGCAGAGAGGCGCAGATCGCAGAGC 1200  
DB 380 GlySerLysArgIleValIlyCysPheAsnCysGlyLysGlnLysIleAlaLysAsn 399  
QY 1201 TCCCGCGCGCCCGCAGAGAGGCGCTGCGAGTGGCGCAGAGAGGCGCAGATGAAG 1260  
DB 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnLysIleGlnMetLys 419  
QY 1261 GACTGACCGCAGCGCGCAGAGCACTTCTGCGCAGATCTGCGCCCGCAGAGAGGCGCGC 1320  
DB 420 AspCysThrGlnArgGlnAlaAsnPheLeuGlnLysIleTrpProSerIlyLysGlyArg 439  
QY 1321 CCGCGCAACTTCTGCGCAGAGCGCAGAGCGCGCGCGCCCGCAGCGTGGCGACCGCGCC 1380  
DB 440 ProGlyAsnPheIleGlnSerArgProGlnProThrAlaProLeuGlnProThrAlaPro 459  
QY 1381 CCGCGCGAGAGCTTCTGCGCTTGAAGAGACACCGCGCGCCCGCAGAGAGCGCGCAGAGC 1440  
DB 460 ProAlaGlnSerPheArgPheGlyGlnThrTrpProAlaProLysGlnGlnProGlnAsp 479  
QY 1441 CCGGAGCGCTTACCGCGAGCGCGCTGACCGCGCGCGCGCTTCTGCGCGCGCGCGCTG 1500  
DB 480 -----ArgGlnProLeuIleSerLeuLysSerLeuPheGlySerAspProLeu 495  
QY 1501 AGCCAG 1506  
DB 496 SerGln 497  
RESULT 10  
Q90ME1 PRELIMINARY; PRT; 492 AA.  
AC Q90ME1; 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrod; Retroviridae; Lentivirus.  
OX NCBI\_taxonomy1676;  
RN (1)  
RP MEDLINE=21565559; PubMed=11709098; DOI=10.1089/08992220152644241;  
RA Engelbrecht S., de Villiers T., Sampson C.C., zur Megede J.,

RA Barnett S.W., van Rensburg E.J.;  
RT "Genetic analysis of the complete gag and env genes of HIV type 1  
RT subtype C primary isolates from South Africa."  
RL AIDS Res. Hum. Retroviruses 17:1533-1547(2001).  
DR EMBL; AF391270; AAK77531.1; -  
DR HSP; P24736; INCD.  
DR GO; GO:0019012; Cytation; IEA.  
DR GO; GO:0003676; P:structural acid binding; IEA.  
DR GO; GO:0005198; P:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag p17; 1.  
DR Pfam; PF00607; Gag p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; ZNF\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.  
KW AIDS; Core protein; Metal-binding; Polyprotein; Zinc; Zinc-finger.  
SQ SEQUENCE 492 AA; 54999 MW; B74F6064D506635D CRC64;  
Alignment Scores:  
Pred. No.: 2,77e-97 Length: 492  
Score: 2437.00 Matches: 457  
Percent Similarity: 94.82% Conservative: 19  
Best Local Similarity: 91.04% Mismatches: 16  
Query Match: 85.81% Indels: 10  
DB: 2 Gaps: 2  
US-09-475-704A-4 (1-1509) x Q90ME1 (1-492)  
QY 1 ATGGGCGCGCGCGCGCAGATCTCTGCGCGCGCAGAGAGTGGACAGATGGAATCCCG 60  
DB 1 MetGlyAlaArgAlaSerIleLeuGlyGlyLysLeuAspThrTrpIlyLysIleArg 20  
QY 61 CTGCGCGCGCGCGCGCAGAGAGCATACATGCTGAGAGCACTGTGTGGCGCGCGCGAG 120  
DB 21 LeuArgProGlyGlyLysLysIleTyrMetLeuLysIleValIleThrAlaSerArgGln 40  
QY 121 CTGAGGCGCTTCTGCGCTGAACCGCGCGCTGTGAGAGCGCGCGCGCTGACAGATGTC 180  
DB 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnGlyCysLysGlnIle 60  
QY 181 ATGAAGCAGCTGACCGCGCGCGCTGAGAGCGCGCAGAGAGCTGCGACCTGTATACAC 240  
DB 61 IleArgGlnLeuGlnProAlaLeuGlnIleThrGlyThrGlnGlnLeuLysSerLeuLys 80  
QY 241 ACCGTGGCAGCGCTTCTGCTGCGTGCAGCGCGCGCATCGAGTTCGCGAGACCGAGAGGCC 300  
DB 81 ThrValAlaThrIleLysCysValIleAlaLysIleGlnValArgAspThrLysAla 100  
QY 301 CTGACAGAGTTCAGAGAGAGAGAGAGAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 101 LeuAspLysIleGlnGlnGlnGlnLysLysCysGlnGlnLysThrGlnGlnIleLysGln 120  
QY 361 GCGGCGCGCAGAGTTCAGAGAGAGAGAGAGAGAGTCCAGAGAGAGAGAGAGAGAGAGAG 420  
DB 121 AlaAspGlyLysIleSerGlnAsnIlyProIleValGlnAsnLysGlnIlyGlnMetVal 140  
QY 421 CACGAGCGCATCAGCGCGCGCGCAGCTGAGAGCGCTGAGAGTGTGAGAGAGAGAGGCC 480  
DB 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnIlyLysAla 160  
QY 481 TTGACCGCGAGAGTATCCCATGTTCAACCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAG 540  
DB 161 PheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlyAlaIleThrProGlnAsp 180  
QY 541 CTGAACAGAGATGTAACAACCGTGGCGCGCGCAGAGCGCGCGCATATGTAAGAGAGC 600  
DB 181 LeuAsnThrMetLeuAsnThrValGlyIlyAsnAlaAlaMetGlnMetLeuLysAsp 200  
QY 601 ACCATCAAGAGAGAGCGCGCGAGTGGAGCGCGCTGACCGCGCGCGAGAGAGAGCGCGCTG 660  
DB 201 ThrIleAsnGlnGlnAlaIleGlnIleThrAspArgIleIleIleProValIleAlaGlyProIle 220

```

QY 661 GCGCCGCGCAGATGGCGACCCCGCGGAGGACATGCGCGCGCCACAGACCTTG 720
Db 221 AATPCTGlyGlnMetArgGlnProArgGlySerAspIleAlaIleYThrThrSerThrLeu 240
QY 721 CAGGACCAATCGCTGATGACCAAGACCCCGCGTGGCGCGGACATCTCAAG 780
Db 241 GlnGlnGlnIleAlaThrMetThrSerAlaProValGlyGlnIleYTrp 260
QY 781 CCGTGATCATCTGGGCTTGAACAAGATCTGGGATGTACAGCCCGTGAGATCTG 840
Db 261 ArgTrpIleIleLeuGlyLeuAlaValArgMetYSerProValSerIleLeu 280
QY 841 GACATCGCGCGGCGCCAGAGGCGCTTCCGAGTACGTAGTGAACCGCTTTCAGACC 900
Db 281 AspIleValGlnGlyProValGlnProValArgPheValAspArgPheLeuYTrp 300
QY 901 CTGCGCGCGCAGGCGCAGGACCAAGACGTGAAGATGTGATGACGAGACCTGTGCTG 960
Db 301 LeuArgAlaGlnGlnIleAlaThrGlnSerValYAsnTrpMetThrAspThrLeuVal 320
QY 961 CAGAACGCCAACCCCGCATCTGCAAGACCATCTGCGCTCTGCGCCCGCGCCCTG 1020
Db 321 GlnAlaAlaAsnProAspCysAspYIleIleLeuArgGlyLeuGlyProGlyAlaThrLeu 340
QY 1021 GAGGATGATGACCGCTGCGCAGGCGGTGGGCGCGCGCCAGGCGCGCGCTG 1080
Db 341 GlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyAlaThrValLeu 360
QY 1081 GCGGAGCGCATGAGCGGCGCAAGCGGTGAACATCATGATGACAGAGCACTTCAAG 1140
Db 361 AlaGlnIleMetSerGlnAlaAsnSerThrAlaIleMetMetGlnArgGlyAsnPheYs 380
QY 1141 GCGCCCGCGCGCAGCTCAAGCTTCAAGCTTCAAGCTGCGGAGAGGCGCAGTCCCAAGAC 1200
Db 381 GlyProValArgAsnIleValCysPheAsnGlyGlyGlnGlyAlaThrValLeu 400
QY 1201 TGCCGCGCGCGCGCAGAGAGGCTGCTGAGAGTGGCGGAGAGGCGCGCAGATGAG 1260
Db 401 CysAlaGlnAlaProAlaArgIleValCysGlyGlyGlnGlyAlaThrValLeu 420
QY 1261 GACTGACCGGAGCGCGCAGGCACTTCTGGGCAAGATCTGGCGCAGGCGCAGGCGCGC 1320
Db 421 AspCysTrpGlnAlaArgGlnAlaAsnPheLeuGlyIleTrpProSerHisIleValArg 440
QY 1321 CCGCGCACTTCTGTCAGAGACCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db 441 ProGlyAsnPheLeuGlnAsnArgThrGln-----ProThrAlaPro 454
QY 1381 CCGCGCGAGAGCTTCTGCTGAGAGAGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
Db 455 ProAlaGlnSerPheAlaArgPheGlnIleThrAsnProAlaProValGlnGlnProValAsp 474
QY 1441 CCGGAGCGCTACCGGAGACCGCTTACCGCGCGCTGCGCAGCGCTGTGGCAGCGCGCGCTG 1500
Db 475 -----ArgGlnProLeuThrSerLeuYSerLeuPheGlySerAspProSer 490
QY 1501 AGCCAG 1506
Db 491 SerGln 492

```

## RESULT 11

```

OQ0N98 PRELIMINARY; PRT; 491 AA.
AC OQ0N98;
DT 01-MAY-2000 (TReMBLrel. 13. Created)
DT 01-MAY-2000 (TReMBLrel. 13. Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25. Last annotation update)
DE Gag.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Lentivirinae.
OX NCBI_TaxID=11676;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=93IN101;
RX MEDLINE=9433500; PubMed=10505681; DOI=10.1089/089922299310223;
RA Mochizuki N., Otsuka N., Matsuo K., Shino T., Kojima A., Kurata T.,
RA Sakai K., Yamamoto N., Isomura S., Dhole T.N., Takebe Y., Matsuda M.,
RA Tatum M.;
RT "An infectious DNA clone of HIV type 1 subtype C.";
RL AIDS Res. Hum. Retroviruses 15:1321-1324(1999).
DR EMBL; AB023804; BA05225.1; -.
DR HSP; P24736; INCP.
DR GO; GO:0019012; C:viralion; IEA.
DR GO; GO:0003676; F:molecular acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016032; F:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR010999; Retroviral matrix.
DR InterPro; IPR000071; Retroviral p17.
DR InterPro; IPR008916; Retroviral capsid C.
DR InterPro; IPR008919; Retroviral capsid_N.
DR InterPro; IPR001878; Znt_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zF_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIV1MATRIX.
DR SMART; SM00343; ZNF_C2HC; 2.
DR PROSITE; PS0158; ZF_CCHC; 2.
KW AIDS; Core protein; Metal-binding; Polyprotein; Zinc; Zinc-finger.
SQ SEQUENCE 491 AA; 54809 MW; 0B719BF55C771A8B CRC64;

```

## Alignment Scores:

Pred. No.:	3.22e-97	Length:	491
Score:	2435.50	Matches:	463
Percent Similarity:	95.42%	Conservative:	16
Best Local Similarity:	92.23%	Mismatches:	12
Query Match:	85.76%	Indels:	11
DB:	2	Gaps:	3

US-09-475-704a-4 (1-1509) x OQ0N98 (1-491)

```

QY 1 ATGGCGCGCGCGCGCAGCATCTTGGCGCGGAGAGCTGGAACAAGTGGAGAAAGATCGC 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAsnProValArgIleArg 20
QY 61 CTGCGCGCGCGCGCGCAGAGCACTTCAAGCACTGTGTGGCGCAGCGCGAG 120
Db 21 LeuArgProGlyGlyValYsAsnIleValMetLeuYsIleValAlaThrAlaSerArgGln 40
QY 121 CTGAGAGGCTTGGCTGGAACCGCGCGCTGTGGAACCGCGGAGGCTGCAAGCATATC 180
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrAlaGlnGlyCysValGlnIle 60
QY 181 ATGAAGCAGTGCAGCGCGCGCTGAGACCGGCAACCGAGAGGCTGCGACCTGTACAC 240
Db 61 IleValGlnLeuGlnProAlaLeuGlnThrGlyThrGlnIleLeuYsSerLeuValAsn 80
QY 241 ACCGTGGCCACCTGTCTGCTGCGAGACCGCGGATGAGGTCCGCGACACCAAGAGGCC 300
Db 81 ThrValAlaThrLeuYrCysValHisAlaGlyIleIleValAlaArgAspThrValAla 100
QY 301 CTGGAACAAGTGCAGAGAGAGAGCAACAAGTCCAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 101 LeuAspValIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
QY 361 GCGGAGCGCAGAGTGAACGAACTACCCCATCTGTGCAAGACTGCGAGGCGCAGATGTG 420
Db 121 AlaAspGlyValSerGlnAsnYrProIleValGlnAsnLeuGlnGlnGlnGlnGlnGln 140
QY 421 CACGAGGCATCAGCGCGCGCGCGCAGGAGCGCTGAGAGTGAAGTGAAGTGAAGAGGCC 480
Db 141 HisGlnAlaIleSerProAlaGlnThrLeuAlaAlaThrValValIleGlnGlnValAla 160
QY 481 TTCAGCGCGCGAGGTGATCCCATGTTCAACCGCGCTGAGCGAGGCGCGCACCGCGCGAG 540

```

```

Db      161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaLeuThrProGlnAsp 180
Qy      541 CTGAACACGAGTTGAACACCGTGGCGGCGCACAGCGCCGATGACATGCTGAAGAC 600
Db      181 LeuAenThrMetLeuAenThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLysAsp 200
Qy      601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCGCGCTGCACCGCGCTGACGCGCGCGCG 660
Db      201 ThrIleAenGluGlnAlaIleGluThrPheArgLeuHisProIleHisAlaGlyProIle 220
Qy      661 GCCCGCGCGAGATGCGGACCGCGCGCGAGCATGCGCGCGCGCGCACAGACCTG 720
Db      221 AlaProGlyGlnMetArgGluProArgLysSerAlaIleGlyThrThreSerLeu 240
Qy      721 CAGGACGAGATGCTGTGATGACAGCAACCCCGCGTGCCTGGCGGACATCTACAG 780
Db      241 GlnGluGlnIleAlaIleThrMetThrGlyAenProProValGlyAspIleTyrLys 260
Qy      781 CGGTGGATCATCTGGGCTGAAACAAAGATCTGCGGATGATACGCCCTGAGCATCTG 840
Db      261 ArgTrpIleIleLeuGlyLeuAenLysIleValArgMetTyrSerProValSerIleLeu 280
Qy      841 GACATCCGCGAGGCGCGCGAGGCGCTCCGCGACTAGCTGAGCGCTTCTTCAAGAC 900
Db      281 AspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArgPheLysThr 300
Qy      901 CTGCGCGCGAGAGCGCACCGAGACGTGAAGACTGATGACCGAGACCTGTGTGTG 960
Db      301 LeuArgIleGluGlnAlaIleThrGlnAspValLysAenThrMetThrAspThrLeuVal 320
Qy      961 CAGAACCGCGCAACCGCGATGCAAGACATCTCTGCGCGCTGCGCGCGCGCGCGCG 1020
Db      321 GlnAenAlaAenProAspLysThrIleLeuArgAlaLeuGlyProValAspSerLeu 340
Qy      1021 GAGGAGATGATGACCGCGCTGCGAGGCGCGTGGCGCGCGCGCGCGCACAGGCGCG 1080
Db      341 GlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisIleAlaArgValLeu 360
Qy      1081 GCCGAGCGGATGAGCGAGCGCACAGCGTGAACATCATGATGACAGAGACCACTTCA 1140
Db      361 AlaGlnAlaMetSerGlnAlaAenSer---ThrIleLeuMetGlnArgSerAenPheLys 379
Qy      1141 GGGCGCGCGCGCAAGCTCAAGTCTTCACTGCGCGAGAGAGGCGCACATGCCCAAGAC 1200
Db      380 GlySerLysArgIleValLysCysPheAenCysGlyLysGluGlyHisIleAlaLysAen 399
Qy      1201 TGCGCGCGCGCGCGCAAGAGGCGCTGGAAGTGGCGGCAAGAGGCGCACAGATGAAG 1260
Db      400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleMetLys 419
Qy      1261 GACTGACACGAGCGCGAGCGCAACTTCTGGGCAAGATCTGGCGCACAGCAAGGCGCG 1320
Db      420 AspCysThrGlnArgGlnAlaAenPheLeuGlyLysIleTrpProSerHisIleGlyArg 439
Qy      1321 CCCGCGCACTTCTGCAAGACCGCGAGCGAGCGCGCGCGCGCGCACCGTGCACCG 1380
Db      440 ProGlyAenPheLeuGlnSerArgProGlu-----ProThrAlaPro 453
Qy      1381 CCCGCGAGAGCTTCCGCTTGAAGAGACCAACCGCGCGCGCGCGCACAGAGAGCG 1440
Db      454 ProAlaGlnSerPheArgPheGlnGluGlnThrTrpProAlaProLysGlnLysProLysAsp 473
Qy      1441 CGCGAGCCCTACCGGAGCGCGCTGACCGCGCTGCGACAGCGCTTGTGCGAGCGCG 1500
Db      474 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspLeu 489
Qy      1501 AGCCAG 1506
Db      490 SerGln 491

```

RESULT 12  
Q901X0

```

ID      Q901X0      PRELIMINARY;      PRT;      492 AA.
AC      Q901X0:
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Gag protein.
GN      Name=gag;
OS      Human immunodeficiency virus 1.
OC      Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX      NCBI_TaxID=11676;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21565558; Pubmed=11709097; DOI=10.1089/0889220152644232;
RA      van Hammele J., Williamson C., Kim B., Morris L., Carr J.,
RT      "Characterization of full-length HIV type 1 subtype C sequences from
RT      South Africa."
RN      [2]
RP      AIDS Res. Hum. Retroviruses 17:1527-1531(2001).
RA      van Hammele J.H., Williamson C., Kim B., Morris L., Carr J.,
RA      Abdool Karim S.S., McCutchan F.
RL      Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY043176; AAL05334.1; -.
DR      HSBP; 070622; 1HYN.
DR      GO; GO:0019012; C:viralion; IEA.
DR      GO; GO:0003676; F:nucleic acid binding; IEA.
DR      GO; GO:0005198; F:structural molecule activity; IEA.
DR      GO; GO:0016032; P:viral life cycle; IEA.
DR      Pfam; PF00540; Gag_D17; 1.
DR      Pfam; PF00607; Gag_D24; 1.
DR      Pfam; PF00098; zF-CCHC; 2.
DR      PRINTS; PR00939; C2HCZNFINGER.
DR      PRINTS; PR00343; ZNF_C2HC; 2.
DR      SMART; SM00343; ZNF_C2HC; 2.
DR      PROSITE; PS0158; zF_CCHC; 2.
KW      AIDS; Core protein; Metal-binding; Polyprotein; zinc; Zinc-finger.
SQ      SEQUENCE 492 AA; 55075 MW; 11966A2F3D272ED0 CRC64;

Alignment Scores:
Pred. No.:      4,12e-97      Length:      492
Score:      2433.00      Matches:      461
Best Local Similarity:      94.62%      Conservative:      14
Best Local Similarity:      91.83%      Mismatches:      17
Query Match:      85.67%      Indels:      10
DB:      2      Gaps:      2

US-09-475-704A-4 (1-1509) x Q901X0 (1-492)
Qy      1 ATGGCGCGCGCGCGCGAGATCTCTGCGCGCGAGAGCAAGTGGAGAGATCCGC 60
Db      1 MetGlyAlaArgAlaSerIleLeuArgGlyLysLysLeuAspLysTrpGlnLysIleArg 20
Qy      61 CTGCGCGCGCGCGCGAGACATCAATGCTGAAGCATGCTGTGGCGCGCGCGAG 120
Db      21 LeuArgProGlyGlyLysLysCysTyrMetLeuLysHisIleIleTrpAlaSerArgGlu 40
Qy      121 CTGAGGCGCTTCCGCTTGAACCCCGCGCTGTGAGACCGCGGAGGCGTGAAGCATC 180
Db      41 LeuGlnArgPheAlaLeuAenProGlyLeuLeuGluThrLysGluGlyCysIleGlnIle 60
Qy      181 ATGAAGACGCTGACCGCGCGCTGACACCGCGCACCGAGAGCTGCGAGCTGTACAC 240
Db      61 IleAenGlnLeuHisProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuLysAen 80
Qy      241 ACCGTGGCACCCCTGACTGCTGTGACCGCGCATCGAGTCCGCGACCAAGAGAGGCC 300
Db      81 ThrValAlaThrLeuTyrCysValHisAlaGluIleGluValArgAspThrLysGlnAla 100
Qy      301 CTGACAAAGATCGAGAGAGACAGAAAGTCCACAGCAAGAACCCAGCGCGAGAGAG 360
Db      101 LeuAspLysIleGluGluGlnGlnAenMetIleGlnGlnLysThrGlnGlnAlaLysAla 120

```



QY	361	GGCGAAGGAGATGAGACCAATTCCCATGTGTGCAGAACTGCAGAGGCGCATGTGTG	420
Db	121	ATAAAGGTLyValaSerGlnsnTyProIleValaGlnsnleuGlnGlnMetVal	140
QY	421	SACCAAGGCATGAGCGCGCGCACTCGAAGCGCTGTGGATGAAGGTATGCAAGAAAGGC	480
Db	141	HSglnAlaIleSerProArgThrIleuAlaTrpValaIleGlnGlnValaA	160
QY	481	TTTCAGCGCGGAGGTGATCCCATGTTCACTGCGCTGAGCGAGGAGCGCGCACTCC	540
Db	161	PheSerProGlnValaIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnIle	180
QY	541	CTGAACACAGATGTTGAACACCGGGGGCGCGCAACAGCGCGCACTGCAAGATGCTGAAGAC	600
Db	181	LeuAlaThrMetIleuAlaThrValaGlnGlyAlaValaIleAlaMetGlnMetLeuVala	200
QY	601	ACCATCAACGAGAGGCGCGCGAGTGGAGCGCGCTTGCACTCCGATGAGCGCGCGCGG	660
Db	201	ThrIleAsnGlnGlnAlaIleGlnTrpArgTrpAlaGlnHisProValHisAlaGlyProVal	220
QY	661	GGCGCGCGCGAGATCGCGCACTCCCGCGCGAGACATGCGCGCGCGCGCACTGACCTG	720
Db	221	AlaProGlnGlnMetArgGlnProArgGlySerMetIleAlaGlyThrThrSerThrIleu	240
QY	721	CAGAGACAGATGCGCTGTGATGACACAGCACTCCCGCGCTGCGCGCGCGCACTTCAAGAC	780
Db	241	GlnGlnGlnIleAlaTrpMetThrIleAlaMetProProIleProValaGlyAlaIleTyTrp	260
QY	781	CGGTGGATCATCTCTGGCGCTGAACAAGATCGTGCGATGTACAGCGCGCGTGAACATCTG	840
Db	261	ArgTrpIleIleLeuGlnIleuAlaValaIleValaGlnMetTyrSerProValSerIleLeu	280
QY	841	GACATCCGCGAGGCGCGCAAGAGCGCTTCCGCGCACTGACAGCGCGCTTCAAGAC	900
Db	281	AspIleArgGlnGlnIleProIleGlnIleProIleAlaGlnArgTyrValaAspArgPheIleu	300
QY	901	CTGGCGCGCGAGCGCGCACTCGAGAGCTGAAGATCGATGATCCAGACCGCTGTGTG	960
Db	301	LeuArgAlaGlnGlnIleAlaThrGlnMetValaIleuAlaValaIleTrpMetThrGlnThrIleuVala	320
QY	961	CAGAACGCGCACTCCGCACTGCAAGACCAATCCGCGCGCTGCGCGCGCGCGCGCACTG	1020
Db	321	GlnsnAlaAlaMetSerGlnAlaAlaMetIleAlaValaIleMetGlnIleuSerAlaPheIleu	340
QY	1021	GAGAGATGATGACCGCGCTGCGAGAGCGCTGGCGCGCGCGCGCGCAAGCGCGCGCTGTG	1080
Db	341	GlnGlnMetMetThrAlaCysGlnGlyAlaGlnIleGlyProGlnIleValaAlaGlnIleu	360
QY	1081	GGCGAGGCGAGAGCGCGCGCAAGCGCTGAACATGATGATGACAGAAAGCACTTCAAG	1140
Db	361	AlaGlnIleAlaMetSerGlnAlaAlaMetIleAlaValaIleMetGlnIleuSerAlaPheIleu	380
QY	1141	GGCGCGCGCGCACTCAAGTGTTCACATGCGCGCGCAAGAGCGCGCACTGCGCGCAAGAC	1200
Db	381	GlyProIleAlaGlnIleValaIleCysPheAsnIleCysGlnIleGlnIleAlaAlaGln	400
QY	1201	TGCGCGCGCGCGCGCAAGAGCGCTGTGGAAGTGCAGCAAGAGCGCGCACTCAATGAAG	1260
Db	401	CysArgTrpAlaProAlaGlnIleValaGlnIleCysTrpIleCysGlnIleGlnIleGlnIleMetIleu	420
QY	1261	GACTGCACGAGCGCGCAAGCGCACTCTGTGGCAAGATCTGAGCGCGCAACAAGCGCGC	1320
Db	421	AspCysIleIleGlnArgGlnAlaAlaPheIleuGlnIleIleTrpProSerAlaValaArg	440
QY	1321	CCGCGCACTTCTGTGCAAAACCGCAGGAGCGCGCGCGCGCGCACTGCGCGCGCGCGC	1380
Db	441	ProGlnIleAsnIleLeuGlnIleAlaMetArgProGln-----ProThrAlaPro	454
QY	1381	CCGCGCGAGAGCTTCCGCTGAGAGAAACAACCGCGCGCGCGCGCGCAAGCGCGCAAGAC	1440
Db	455	ProAlaGlnIleSerPheIleuPheGlnGlnIleThrProAlaProIleuGlnIleThrIleVala	474
QY	1441	CGGAGCGCTTACCGCGAGCGCGCTTGACCGCGCTGTGCGAGCGCGCGCGCTG	1500

Accession	Protein	Length	Score	E-value
Db	Arginylproline-histidine-leucine-serine-leucine-phenylglycine-aspartate	475	100.0	0.0
Qy	AgcCag	1501	100.0	0.0
Db	SerGln	491	100.0	0.0

## RESULT 13

ID	Q7SLX7	PRELIMINARY;	PRT;	491 AA.
AC	Q7SLX7;			
DT	01-OCT-2003	(TrEMBLrel. 25, Created)		
PT	01-OCT-2003	(TrEMBLrel. 25, Last sequence update)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE	Gag protein.			
GN	Name=gag.			
OS	Human immunodeficiency virus 1.			
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus			
OX	NCBI_Taxid=11676;			

RN  
 [1]  
 RP  
 SEQUENCE FROM N. A.  
 RA  
 Gupta R. M., Praead V., Singh B., Seth P.,  
 RL  
 Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases

DR EMBL; AF533131; AAP87971.1; -.  
DR HSPB; Q70622; IHN.  
DR GO; GO:0019012; C:Vtition; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:ribal life cycle; IEA.

DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR010999; Retrovir\_matrix.

DR InterPro: IPR0000071; Retrovir\_p17.  
DR InterPro: IPR008916; Retrov\_capsid\_C.  
DR InterPro: IPR008919; Retrov\_capsid\_N.

DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gaa p17; 1.

DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; zf-CCHC; 2.

DR SMART; SM00343; ZnF C2HC; 2.  
DR PROSITE; PS50158; ZF CCHC; 2.  
AIDS Conserved: Bel

FM	AIDS, core protein; polyprotein.
SO	SEQUENCE 491 AA; 54815 MW; 36CD9A56CAEFAD7C CRC64;

**Alignment Scores:**

Pred. No.:	5,28e-97	Length:	421
Score:	2430.50	Matches:	463
Percent Similarity:	94.82%	Conservative:	13
Best Local Similarity:	92.23%	Mismatches:	15
Query Match:	85.58%	Indels:	11
DB:	2	Gaps:	3

US-09-475-704A-4 (1-1509) x Q7SLX7 (1-491)

QY 1 ATGGGCGCCGCGCAGCATCTTGGCGCGGAGAGCTGCAAGTGGAGAAGATCCGC 60

Db  
1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysIrrGluIuysIleArg 20

61 CTGCGCCCGCGCGCAAGAACTACATGCTGAAGCACCTGTGTGGCCAGCCCGGAG 120

21 leuArgProGlycylLysLysValHisTyrMetLeuLysHisValAlaIleArgGln 40

121 TTGAGGACCTTCGCCCCGAAACCCCGGCGCTGGAGACCCGCGAGGACCTGCAAGCAGATC 188

181 ATGAAGAGCTGCAAGCCGCCCTTCAGACACGGAGAGCTGGCAGCCTGTACAC 244

Db 61 LeuLyselInLeuGlProAlaLeuGlInThrGlyThrGluGluLeuArgSerLeuTyrAsn 80

241 ACCGTGGCCACCCCTGTACTGCGTCACGCCGGGCATCGAGGTCGGCACCAACGAGGCGC 300

Db 81 ThrValAlaThrLeuTYrCysValHisIaGlyIleGluValArgAspThrIysGluAla 1000



QY	301	CTGGAAGAAGTGAAGAGGAGGAGAAACAAGTCCGACGAGAAACACCGAGGCGCAAGAG	360
DB	101	LeuAaRyLlLeGIuGtUeGtImbnyLvgLgInGnLlyeThGtImbLlaLyuLvgLy	120
QY	361	GCGGAGCGCAAGTGAAGCGCAAGAACTAACCCCATCGTGCAGAACCTTGACGGCCAGATGCTG	420
DB	121	AlaAaRgLyLvgValSerGlnmbnThyProIleValGlnmbLueGlnGlyImbVal	140
QY	421	CACCAAGGCGATGAGCCCCCGCAACCCCTGAACGCGCTGGGTGAAGTATGAGAGAAAGCC	480
DB	141	HsGlnAlaIleSerProGhThrLeuambLAtRvAllyeValIlLeGluGlnyAla	160
QY	481	TTGACGCGCGAGGTATCCGCAATGTTCAACCGGCGCTGAGGAGGGGCGCAACCCCGCAAGAC	540
DB	161	PheSerProGlnValIlLeProMetPheThrAlaLeuSerGlnGlyAlaThrProGlnmb	180
QY	541	CTGAACACAGATGTTGAACAACCGTGGGGGCGGCAACAGGCGCGCATGACAGATGTGAAGAC	600
DB	181	LeuambThrMetLeuambThrValGlyGlnHsGlnAlaIleMetGlnmetLeuAaR	200
QY	601	ACCATCAACAGAGAGAGCGCGCGAGTGGGACGCGCTTGCAACCCCGTCAAGCGCGGCGCGTG	660
DB	201	ThrIleasnGlnGlnAlaIleGluThrAaRgLeuHsLeuProValHsAlaGlyProIle	220
QY	661	GCCCCCGGCGAGTGGGCGACCCCGCGGCGACGACATCGCGGGCGCAACGCAACCGCTG	720
DB	221	AlaProGlyGlnMetAaRgInProAaRgLySerAaRlLeaGlyThrThSerThxleu	240
QY	721	CAGAGACAGATGCGCTGGATGACACAAACCCCGCGTGGCGCGTGGGCGACATCTTAACAG	780
DB	241	GlnGlnGlnIleAlaThrMetThxmbnProProValGlyAaRlLeuThx	260
QY	781	CGGTGATCATCTTGAGGCGCTGAACAAGATCGTGCAGATGTACAGCCCGGTGACATCTTG	840
DB	261	ArgTrpIleIleLeuGlyLeuambLyLleValAaRmetThxSerProValSerIleleu	280
QY	841	GACATCCGCGCAGGGCGCCAGAGGCGCTTCGCGCACTAGTGAAGACGCTTCTTCAAGAC	900
DB	281	AspIleLyLgInGlyProLyGlnProAaRgAaRyTValAaRgThrPheThx	300
QY	901	CTGCGCGCGGAGCAGGCGCAACGAGAACGTAAGATCAAGACCGACCGCTGGTG	960
DB	301	LeuAaRgAlaGlnGlnAlaThrGlnmbRAllyAaRmbThrMetThxAaRlLeuAaI	320
QY	961	CAGAACGCGCAACCCCGACTGCAAGACCAATCTCGCGCGCTTCGACCGCGCGCGCAACCTG	1020
DB	321	GlnmbAlaAaRProAaRyThrIleAaRgAlaLeuGlyProGlyAlaThxleu	340
QY	1021	GAGGAGATGATGACCGCTGCCAGGGCGGTGGCGGCGCGCGCAACAGCGCGCTGCTG	1080
DB	341	GlnGlnmbMetThrAlaCyLeGlnGlyAlaGlyLgProSerHsLyAlaAaRyAlleu	360
QY	1081	GCGGAGGAGATGAGCGCAGGCGCAACAGCGTAACATCATGATGACAAAGGAACTTCAAG	1140
DB	361	AlaGlnAlaMetSerGlnThxLySer---ThrIleMetMetGlnAaRSerHsmbLy	379
QY	1141	GGCGCCCGCGCGCAAGTCAAGTGTCTTCAATCTGCGCAAGAGGCGCAATGCGCAAGAC	1200
DB	380	GlySerLyAaRgIleValLyLeuAaRmbnCyGlyLvgGlnGlnHsIleAlaAaRgmb	399
QY	1201	TGCGCGCGCGCGCGCAAGAGGCGCTGTAAGTGCAGGCGCGCAAGAGGCGCAACATGAAG	1260
DB	400	CyAaRgAlaProAaRgLyLvgLySerThxLyCyGlyLvgGlnGlnHsIleGlnmetLy	419
QY	1261	GACGTGACACGAGCGCGCAGGCGCAACTTCGCGGCAAGATGTGGCGCGCAACAGGCGCGC	1320
DB	420	AaRyThrGlnAaRgGlnAlaAaRmbHsLeuGlyLvgIleThrProSerHsLyAaRyAaRg	439
QY	1321	CCCGGCAACTTCTTGCAAGACCGCAAGCGACCGCGCGCGCGCAACGCTGCGCAACCGCGCC	1380
DB	440	ProGlnAaRmbHsLeuGlnSerThxProGln-----ProThxAlaAaR	453
QY	1381	CCCGCGGAGAGCTTCGCTTCGAGGAGACCAACCCCGCGCGCGCGCGCAAGAGCGCGCAAGAC	1440

```

Db      454 PFOAIAgIuSerPheargrhegiuInuItrthPFOAIAProLySgInuIuLeuLYAsp 473|||||
Qy      1441 CGCGAGCCCTTACCGGAGGCCCTTGACCGGCTTGCGAGCCTGTTGGCAGCGGCCCTTG 150|||||
Db      474 -----ArgIuProLeuItrhSerLeuLYSerLeuhegiSerAspProLeu 485|||||
Qy      1501 AGCCAG 1506
Db      490 Sergin 491
      |||||
RESULT 14
Q90127 PRELIMINARY; PRT; 492 AA.
ID      Q90127
AC      Q90127;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT      01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE      Gag protein.
GN      Name=gag;
OS      Human immunodeficiency virus 1.
OC      Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX      NCBI_Taxid=11676;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=21565558; PubMed=11709097; DOI=10.1089/0889220152644232;
RA      van HammeIen J., Williamson C., Kim B., Morris L., Carr J.,
RA      Abdool Karim S.S., McCutchan F.;
RT      "Characterization of full-length HIV type 1 subtype C sequences from
RL      South Africa." ;
RL      AIDS Res. Hum. Retroviruses 17:1527-1531(2001).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      van HammeIen J.H., Williamson C., Kim B., Morris L., Carr J.,
RA      Abdool Karim S.S., McCutchan F.;
RL      Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL, AY0431173; MAF05307.1; -.
DR      HSP; P24736; INCP.
DR      GO: GO:0019012; C:virlon; IEA.
DR      GO: GO:0003676; F:nucleic acid binding; IEA.
DR      GO: GO:0005198; F:structural molecule activity; IEA.
DR      GO: GO:0016032; P:viral life cycle; IEA.
DR      Pfam; PF005407; Gag_p17; 1.
DR      Pfam; PF00607; Gag_p24; 1.
DR      Pfam; PF00098; zfcCHC; 2.
DR      SMART; SM00343; zNF_C2HC; 2.
DR      PROSITE; PS50158; zF_CCHC; 2.
KM      AIDS; Core protein; Folyprotein.
SQ      SEQUENCE 492 AA; 55205 MW; 55CB9BE2D591BA3A CRC64;
Alignment Scores:
Pred. No.: 5.55e-97 Length: 492
Score: 2430.00 Matches: 456
Percent Similarity: 94.82% Conservative: 18
Best Local Similarity: 91.24% Mismatches: 16
Query Match: 85.56% Indels: 10
DB: 2 Gaps: 2
US-09-475-704A-4 (1-1509) x Q90127 (1-492)
Qy      1 ATGGGCGCCCGCGCAGCATCTCTGGCGGGCGAGAGCTGACAGAGTGCGAGAAAGATCCGC 60
Db      1 MetGyIAaYgAlaSerIleLeuArGIGlyIuLYSerLeuAspThrTrpGIuYsIleArg 20
Qy      61 CTGGCGCCCGCGGGGGAAGAGACTATAGCTGAGAGCACTGTGTGGGCGAGCGCGAG 120
Db      21 LeuAaGrProGIyGIuLYsYbScYsTyTrMetLeuLYShsIleValTrpAlaSerArgIu 40
Qy      121 CTGGAGGAGCTTGCAGCTGAACCCCGGCTGTCTGAGAGACCGCGAGGCTGCAAGCAGATC 180
Db      41 LeuGIuArGpHeSerLeuAsnProGIyLeuLeuGIuItrhSerGIuIuLYsYbSgInuIle 60
Qy      181 ATGAAGCAGCTTGACCGCGCCCTGCAAGACCGGACCGAGAGAGCTGCGCAGCTGTACAC 240

```

Db 61 MetLeuGlnLeuInProAlaLeuGlnThrGluGluLeuSerLeuYrAsn 80  
Qy 241 ACCGTGGCCACCTGTACTGCTGTCAGCGCGGATCGAGTCCGGACACCAAGAGGCC 300  
Db 81 ThrValAlaThrLeuPheCysValHisGluYrLeuAlaValArgAspThrLysGlnAla 100  
Qy 301 CTGGACAAGATGAGAGGAGGAGGAGCAAGTCCACACAGAACCCAGAGGCCAGAGAG 360  
Db 101 LeuAspLysLeuGluGluGlnGlnAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120  
Qy 361 GCCGAGCGGAAAGGTGAGCAAGAACTACCCCATGTGTCAGAACTCGAGGCGCAGATGATG 420  
Db 121 AlaAspGlyThrValSerGlnAsnYrProIleValGlnAsnLeuGlnGlnMetVal 140  
Qy 421 CACCGAGCGCATGACCGCCCGGACCCCTGAGACGCTGGGTGAGGTGATCGAGAGAGGCC 480  
Db 141 HisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLysAla 160  
Qy 481 TTCAGCCCGGAGGTGATCCCATGTTACCGCCCTGAGGAGGAGGCGCACCCCGACAGAC 540  
Db 161 PheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGlnAsp 180  
Qy 541 CTGAACACCATGTTGAACACCGTGGCGGCGCACAGCGCGCATGAGATGCTGAAGAGAC 600  
Db 181 LeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysAsp 200  
Qy 601 ACCATCAAGAGAGGCGCGCGGAGTGGACCGCTGTGACCCCGTGTGACGCGCGCCGTG 660  
Db 201 ThrIleAsnGluGlnAlaIleGluTrpAspArgLysHisProValHisIleAlaGlyProIle 220  
Qy 661 GCCCGCGCGCAGATGCGCGCACCCCGCGGAGGAGCATGCGCGGCGCGCACAGACCCGTG 720  
Db 221 AlaProGlyGluMetArgLysProArgLysSerAspIleAlaGlyThrThrSerThrLeu 240  
Qy 721 CAGAGACGATGCTGCTGATGACAGCAACCCCGCTGCGCGTGGCGGACATCTTACAG 780  
Db 241 GlnGluGlnIleAlaTrpMetThrAsnAsnProProValProValGlyAspIleThrLys 260  
Qy 781 CGGTGGATCATCTGGGCGCTGAACAAGATCGCGCGGATGTACAGCCCGTGAAGCATCTCTG 840  
Db 261 ArgTrpIleIleLeuGlyLeuAsnLysIleValAlaGlyMetYrSerProValSerIleLeu 280  
Qy 841 GACATCCGCGCAGAGGCCCGCAAGAGGCTTCCGCGACTACGTGAGACCGCTTCTTCAAGAC 900  
Db 281 AspIleArgGlnGlyProLysGluProLysArgAspYrValAspArgPhePheLysThr 300  
Qy 901 CTGCGCGCGCGAGGAGGCCACCGAGCATGAAAGACTGATGACCGAGACCCCTGTGTGTG 960  
Db 301 LeuValGalaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuVal 320  
Qy 961 CAGAAAGCCAAACCCGACTGCAAGACCAATCCGCGCGCTCGGCGCGGCGCGCACCCCTG 1020  
Db 321 GlnAsnAlaAsnProAspCysLysThrIleLeuAlaGlnAlaGlyProGlyAlaThrLeu 340  
Qy 1021 GAGAGATGATGACCGCTGCGAGGCGTGGCGGCGCCCGGCGACAGCGCGCGCTGTG 1080  
Db 341 GluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysValAlaValLeu 360  
Qy 1081 GCCAGAGGATGAGGAGGCCCAACGCTGAACATCATGATGACAGAGAGCACTTCAAG 1140  
Db 361 AlaGluAlaMetSerGlnThrAsnAsnAlaAsnIleMetGlnArgSerAspPheLys 380  
Qy 1141 GGGCCCGCGGCAACGTCAAGTGTCTTCACTGCGGCAAGAGGCGCACATGCGGCAAGAC 1200  
Db 381 GlyProArgArgIleIleLysCysPheAsnLysGlyLysGluGlyHisLysAlaArgAsn 400  
Qy 1201 TGGCGCGCGCGCGAGAGAGGCTGTGAAGTGTGGCGAGAGGCGCACAGATGAG 1260  
Db 401 CysArgAlaProArgLysGlyGlyCysThrLysCysGlyLysGluGlyHisGlnMetLys 420  
Qy 1261 GACTGCAACGAGCGCGAGGCCAACCTTCTGGGCAAGATCTGGCCGACACAGAGCGCC 1320

Db 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440  
Qy 1321 CCCGCAACTTCTCTGACAGAACCGGAGCGCGCGCCCGCCACCGCTGCGACCGCGCC 1380  
Db 441 ProGlyAsnPheLeuGlnAsnArgProGlu-----ProThrAlaPro 454  
Qy 1381 CCCGCGAGAGCTTCCGCTTGAAGAGACACCCCGCGCCCGCAAGCAGAGCCCAAGAGAC 1440  
Db 455 ProAlaGluSerPheArgPheGluGluThrThrProAlaProLysGlnGluProArgGlu 474  
Qy 1441 CGCGAGCCCTACCGGAGCGCCCTGACCGCGCTGCGAGCGCTGTTCGAGCGCGCCCTG 1500  
Db 475 -----ArgGluProLeuThrSerLeuLysSerLeuPheGlySerAspProLeu 490  
Qy 1501 AGCCAG 1506  
Db 491 SerGln 492  
RESULT 15  
ID Q90ME2 PRELIMINARY; PRT; 492 AA.  
AC Q90ME2;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21565559; PubMed=11709096; DOI=10.1089/08892220152644241;  
RA Engelbrecht S., de Villiers T., Sampson C.C., zur Megede J.,  
Barnett S.W., Van Rensburg E.J.;  
RT "Genetic analysis of the complete gag and env genes of HIV type 1  
subtype C primary isolates from South Africa";  
RL AIDS Res. Hum. Retroviruses 17:1533-1547(2001).  
DR BML; AF31269; AK77350.1; -.  
DR HSBP; P24736; INCP.  
DR GO; GO:0019012; C:viralon; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; Zf-CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50158; Zf\_CCHC; 2.  
KW AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 492 AA; 55186 MW; E8023DB89216F43D CRC64;  
Alignment Scores:  
Pred. No.: 5.55e-97 Length: 492  
Score: 2430.00 Matches: 456  
Percent Similarity: 94.82% Conservative: 20  
Best local Similarity: 90.84% Mismatches: 16  
Query Match: 85.56% Indels: 10  
DB: 2 Gaps: 2  
US-09-475-704A-4 (1-1509) x Q90ME2 (1-492)  
Qy 1 ATGGGCGCGCGCGCGCGCATCTGCGCGCGGAGAGACTGGAACAAGTGGAGAAATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspThrTrpGluYrIleArg 20  
Qy 61 CTGCGCGCGCGCGCGAGAGCACTACATGCTGAACACACTGGTGTGGCGAGCGCGAG 120  
Db 21 LeuArgProGlyLysLysHisYrMetLeuLysHisLeuValTrpAlaSerArgGln 40  
Qy 121 CTGAGAGCGTTCGCGCTGAACCCCGCGCTGTGAGACCGCGGCGGTGCAAGCATC 180  
Db 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60



THIS PAGE BLANK (USPTO)



CC assemble into non-infectious virus-like particles which are used as a  
 CC matrix for the proper presentation of an antigen entrapped or associated  
 CC to the immune system of the host  
 XX

5Q Sequence 1509 BP; 320 A; 556 C; 472 G; 161 T; 0 U; 0 Other;

Query Match 100.0%; Score 1509; DB 3; Length 1509;

Best Local Similarity 100.0%; Pred. No. 5.3e-177;  
 Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGGGCGCCCGCCGACATCTTCCGCGGAGAGAGCTGGAACAATGGAGAGATCCG 60
DB 1 ATGGGCGCCCGCCGACATCTTCCGCGGAGAGAGCTGGAACAATGGAGAGATCCG 60
QY 61 CTGGGCGCCCGCCGACATCTTCCGCGGAGAGAGCTGGAACAATGGAGAGATCCG 120
DB 61 CTGGGCGCCCGCCGACATCTTCCGCGGAGAGAGCTGGAACAATGGAGAGATCCG 120
QY 121 CTGGAGGCTTCCGCTGAAACCCCGGCTGCTGAGACGCGAGAGGCTGACAGATC 180
DB 121 CTGGAGGCTTCCGCTGAAACCCCGGCTGCTGAGACGCGAGAGGCTGACAGATC 180
QY 181 ATGAAGCATGTGAGCCCGCTGCAAGCCGCACTGAGAGAGCTGCGACGCTGTACAAC 240
DB 181 ATGAAGCATGTGAGCCCGCTGCAAGCCGCACTGAGAGAGCTGCGACGCTGTACAAC 240
QY 241 ACCGTGGCCACCTGTACTGCGTGAAGCGCGGCGATGAGGTCCGCGACCAAGAGAGCC 300
DB 241 ACCGTGGCCACCTGTACTGCGTGAAGCGCGGCGATGAGGTCCGCGACCAAGAGAGCC 300
QY 301 CTGGACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 301 CTGGACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 361 GCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 361 GCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 421 CACGAGGCGATGAGCCCGGCAACCTGAAAGCGCTGGAGTGAAGTATCGAGAGAGAGCC 480
DB 421 CACGAGGCGATGAGCCCGGCAACCTGAAAGCGCTGGAGTGAAGTATCGAGAGAGAGCC 480
QY 481 TTCAAGCCCGAGAGTATCCCATGTTCAACGCGCTGAGAGAGAGAGAGAGAGAGAG 540
DB 481 TTCAAGCCCGAGAGTATCCCATGTTCAACGCGCTGAGAGAGAGAGAGAGAGAGAG 540
QY 541 CTGAACAGAGATGTTGAACAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 541 CTGAACAGAGATGTTGAACAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 601 ACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 ACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 GCGCCCGGCGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 GCGCCCGGCGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 CAGGAGAGATGCGCTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 CAGGAGAGATGCGCTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 CCGTGGATCATCTGGGCTGAAACAAGATCGTGGAGATGACAGCCCGGTAGAGATCTTG 840
DB 781 CCGTGGATCATCTGGGCTGAAACAAGATCGTGGAGATGACAGCCCGGTAGAGATCTTG 840
QY 841 GACATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 GACATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 CTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 901 CTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

```

```

QY 961 CAGAACGCCAACCCGACTGCAAGACATCTGCGCGCTTTCGCCCCCGCGCCACCTTG 1020
DB 961 CAGAACGCCAACCCGACTGCAAGACATCTGCGCGCTTTCGCCCCCGCGCCACCTTG 1020
QY 1021 GAGGAGATGATACCGGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 GAGGAGATGATACCGGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 GCCGAGGAGATGAGCCAGGCAACAGCGTGAACATCATGATGACAGAGAGAGAGAGAG 1140
DB 1081 GCCGAGGAGATGAGCCAGGCAACAGCGTGAACATCATGATGACAGAGAGAGAGAGAG 1140
QY 1141 GCGCCCGGCGCAACGTCAAGTCTTCAATGCGGCAAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 GCGCCCGGCGCAACGTCAAGTCTTCAATGCGGCAAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 TCGCGGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 TCGCGGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 GACTGACCGAGCGCGCAAGCACTTCTGAGGCAAGATCTGAGCCAGCCAGAGAGAGAG 1320
DB 1261 GACTGACCGAGCGCGCAAGCACTTCTGAGGCAAGATCTGAGCCAGCCAGAGAGAGAG 1320
QY 1321 CCGGCGCACTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 CCGGCGCACTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 CCGCGCGAGAGCTTCCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 CCGCGCGAGAGCTTCCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 CGGAGGCTTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1441 CGGAGGCTTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 AGCCAGTAA 1509
DB 1501 AGCCAGTAA 1509

```

```

RESULT 2
AAL44549
ID AAL44549 standard; DNA; 1509 BP.
XX
AC AAL44549;
XX
DT 29-AUG-2003 (revised)
DT 08-NOV-2002 (first entry)
XX
DE HIV-1 p55gag polypeptide coding sequence 2.
XX
KW HIV; ds; vaccine; gene; immune response; microparticle;
KW adjuvant surface; poly(alpha-hydroxy acid); poly(hydroxy butyric acid);
KW polyprolactone; poly(orthoester); polycyanocrylate; detergent;
KW submicron emulsion; viral infection; bacterial infection;
KW parasitic infection; HIV-1 p55gag polypeptide.
XX
OS Human immunodeficiency virus 1.
XX
PN MO20226209-A2.
XX
PD 04-APR-2002.
XX
PF 28-SEP-2001; 2001MO-US030540.
XX
PR 28-SEP-2000; 2000US-0236105P.
XX
PR 30-AUG-2001; 2001US-0315905P.
XX
PA (CHIR ) CHIRON CORP.
XX
PI O'hagan D, Otten G, Donnelly JJ, Polo JM, Barnett S, Singh M,

```



XX immunostimulant; gene therapy; gene; ds.  
XX Human immunodeficiency virus; type C.  
OS Synthetic.  
XX WO200204493-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 05-JUL-2001; 2001WO-US021241.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
DR WPI; 2002-154920/20.  
XX  
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
PT applications including DNA immunization or generation of packaging cell  
PT lines, particularly in gene therapy.  
XX  
XX Example 1; Fig 2; 233pp; English.  
PS  
XX The present invention describes expression cassettes comprising a  
XX polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (1). (1) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL3942 to ABL40054 and  
CC ABL06204 to ABL06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 1509 BP; 320 A; 556 C; 472 G; 161 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 1509; DB 6; Length 1509;  
Best Local Similarity 100.0%; Pred. No. 5.3e-177;  
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ATGGGCGCCCGCGCAGCATCTGCGCGGAGAACTGGACAACTGGAGAAATCCGC 60  
Db 1 ATGGGCGCCCGCGCAGCATCTGCGCGGAGAACTGGACAACTGGAGAAATCCGC 60  
  
QY 61 CTGCGCGCCCGCGGCAAGAAAGCACTAATGTGGAAGCACTGTGTGGCCAGCCGCGAG 120  
Db 61 CTGCGCGCCCGCGGCAAGAAAGCACTAATGTGGAAGCACTGTGTGGCCAGCCGCGAG 120  
  
QY 121 CTGAGGGCTTGGCCTGAAACCCCGCCTGTGTGAGAACCGCGAGGGCTGCAAGCAGATC 180  
Db 121 CTGAGGGCTTGGCCTGAAACCCCGCCTGTGTGAGAACCGCGAGGGCTGCAAGCAGATC 180  
  
QY 181 ATGAAGCAGCTGCAAGCCCGCTTGCAACCGGCACTGGAGAGCTGCGACGCTGTACAC 240  
Db 181 ATGAAGCAGCTGCAAGCCCGCTTGCAACCGGCACTGGAGAGCTGCGACGCTGTACAC 240  
  
QY 241 ACCGTGCGCACCTGTATCTGCGTGCACCGCGGCACTGAGTCCGGAGAACCAAGAGGCC 300  
Db 241 ACCGTGCGCACCTGTATCTGCGTGCACCGCGGCACTGAGTCCGGAGAACCAAGAGGCC 300  
  
QY 301 CTGGAAGATCGAGAGAGAGCAAAAGTCCAGCAAGAAAGCCAGCAGGCCAAGAG 360  
Db 301 CTGGAAGATCGAGAGAGAGCAAAAGTCCAGCAAGAAAGCCAGCAGGCCAAGAG 360  
  
QY 361 GCCGACGGCAAGAGAGCAAGTACCCCATCTGTGCAAGAACTGTGCAAGGGCCAGATGTG 420  
Db 361 GCCGACGGCAAGAGAGCAAGTACCCCATCTGTGCAAGAACTGTGCAAGGGCCAGATGTG 420  
  
QY 421 CACGAGGCGATCAGCCCGCGACCTGTGAACGCTGGGTGAAGTATCGAGAGAAAGCC 480

Db 421 CACGAGGCGATCAGCCCGCGACCTGTGAACGCTGTGATGATCGAGAGAAAGCC 480  
QY 481 TTCAGCCCGAGAGTATCCCATGTTCAACCGCCCTGACAGAGGGGCCACCCCGCAGAG 540  
Db 481 TTCAGCCCGAGAGTATCCCATGTTCAACCGCCCTGACAGAGGGGCCACCCCGCAGAG 540  
QY 541 CTGAACAGATGTTGAACAACGCTGGCGGCCACAGCGCGCATATGATCTGAAGAGAC 600  
Db 541 CTGAACAGATGTTGAACAACGCTGGCGGCCACAGCGCGCATATGATCTGAAGAGAC 600  
QY 601 ACCATCAACAGAAAGCCCGCGAGTGGAGACCGCTGTGACCCCGTGCAGCGCCCGCTG 660  
Db 601 ACCATCAACAGAAAGCCCGCGAGTGGAGACCGCTGTGACCCCGTGCAGCGCCCGCTG 660  
QY 661 GCCCGCGCGAGATCGGACCCCGCGGCGAGCATCGCGCGGCGCCACAGACCCCTG 720  
Db 661 GCCCGCGCGAGATCGGACCCCGCGGCGAGCATCGCGCGGCGCCACAGACCCCTG 720  
QY 721 CAGAGCAGATGCTGTGATGACAGCAACCCCGCTGCGGTGGCGCATCTACAG 780  
Db 721 CAGAGCAGATGCTGTGATGACAGCAACCCCGCTGCGGTGGCGCATCTACAG 780  
QY 781 CGGTGATCATCTGGGCTGTGAACAGATGATGAGATGACAGCCCGTGAAGATCTG 840  
Db 781 CGGTGATCATCTGGGCTGTGAACAGATGATGAGATGACAGCCCGTGAAGATCTG 840  
QY 841 GACATTCGCAAGGCGCCCAAGAGACCTTCCGCACTACGTGACCGCTTCTTCAAGAC 900  
Db 841 GACATTCGCAAGGCGCCCAAGAGACCTTCCGCACTACGTGACCGCTTCTTCAAGAC 900  
QY 901 CTGGCGCCCGGAGGCGTCAACCCAGAGATGAAATCTGATACCGAGACCTGTGTG 960  
Db 901 CTGGCGCCCGGAGGCGTCAACCCAGAGATGAAATCTGATACCGAGACCTGTGTG 960  
QY 961 CAGAACGCAACCCCGCATGAGACCATCTGCGCGCTCTCGGCGCCCGGCGCAACCTG 1020  
Db 961 CAGAACGCAACCCCGCATGAGACCATCTGCGCGCTCTCGGCGCCCGGCGCAACCTG 1020  
QY 1021 GAGAGATGATGACCGCTTCCGAGGCGTGGCGGCGCCCGGCGCAAGGCGCGTGTG 1080  
Db 1021 GAGAGATGATGACCGCTTCCGAGGCGTGGCGGCGCCCGGCGCAAGGCGCGTGTG 1080  
QY 1081 GCGAGCGATGAGCGCGCAAGCGTCAAGTCTTCACTGCGGCAAGAGGCCACATCGCAAGAAC 1140  
Db 1081 GCGAGCGATGAGCGCGCAAGCGTCAAGTCTTCACTGCGGCAAGAGGCCACATCGCAAGAAC 1140  
QY 1141 GGCCTCCGCGCAACGTCAGTCTTCACTGCGGCAAGAGGCCACATCGCAAGAAC 1200  
Db 1141 GGCCTCCGCGCAACGTCAGTCTTCACTGCGGCAAGAGGCCACATCGCAAGAAC 1200  
QY 1201 TGGCGGCGCCCGGCAAGAGGGCTGTGGAATGTGGCAAGAGGCCACATCGCAAGAAC 1260  
Db 1201 TGGCGGCGCCCGGCAAGAGGGCTGTGGAATGTGGCAAGAGGCCACATCGCAAGAAC 1260  
QY 1261 GACTGCAACGAGGCGAGGCAACTTCTTGGGCAAGATCTGGCCCAAGAGGCCGCG 1320  
Db 1261 GACTGCAACGAGGCGAGGCAACTTCTTGGGCAAGATCTGGCCCAAGAGGCCGCG 1320  
QY 1321 CCGGCGAATTCTTCTGAGAAACCGGAGCGAGCCCGCGCCCAACGCTGCGCAACCGCC 1380  
Db 1321 CCGGCGAATTCTTCTGAGAAACCGGAGCGAGCCCGCGCCCAACGCTGCGCAACCGCC 1380  
QY 1381 CCGGCGAGAGCTTCCGCTTCAAGAGACCAACCCCGCCCAAGAGAGCCCAAGAGAC 1440  
Db 1381 CCGGCGAGAGCTTCCGCTTCAAGAGACCAACCCCGCCCAAGAGAGCCCAAGAGAC 1440  
QY 1441 CCGAGACCTTACCGGAGCCCTGTACCGGCTGTGCGAGCTTGTGGCAGCGGCCCTGTG 1500  
Db 1441 CCGAGACCTTACCGGAGCCCTGTACCGGCTGTGCGAGCTTGTGGCAGCGGCCCTGTG 1500  
QY 1501 AGCCAGTAA 1509





QY 1261 GACTGACGAGGCGGCAAGCTTCTGGGCAAGATCTGGCCGCAAGGCGCGC 1320  
Db 1261 GACTGACGAGGCGGCAAGCTTCTGGGCAAGATCTGGCCGCAAGGCGCGC 1320  
QY 1321 CCCGCACTTCTTGGAGAACGCGAGCGCGCGCCCAACCTGTGCCACCGCCCC 1380  
Db 1321 CCCGCACTTCTTGGAGAACGCGAGCGCGCGCCCAACCTGTGCCACCGCCCC 1380  
QY 1381 CCCGCGAAGCTTCCGCTTCCAGAGACACCCCCCGCCCAAGAGAGAGCCCAAGAC 1440  
Db 1381 CCCGCGAAGCTTCCGCTTCCAGAGACACCCCCCGCCCAAGAGAGAGCCCAAGAC 1440  
QY 1441 CCGGAGCCCTTACCGGAGCCCTTGAACCGGCTTGGAGGCTTGGCGAGCGCGCCCTG 1500  
Db 1441 CCGGAGCCCTTACCGGAGCCCTTGAACCGGCTTGGAGGCTTGGCGAGCGCGCCCTG 1500  
QY 1501 AGCCAGTAA 1509  
Db 1501 AGCCAGTAA 1509  
RESULT 5  
ABL39958  
ID ABL39958 standard; DNA; 1509 BP.  
AC ABL39958;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Synthetic Gag polynucleotide sequence SEQ ID NO:21.  
XX  
KW Human immunodeficiency virus type C; antigenic HIV type C protein;  
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
KW immunostimulant; gene therapy; gene; ds.  
OS Human immunodeficiency virus; type C.  
OS Synthetic.  
XX  
PN WO200204493-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 05-JUL-2001; 2001WO-US021241.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (CHIR ) CHIRON CORP.  
XX (UIST-) UNIV STELLENBOSCH.  
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
PI MPI; 2002-154920/20.  
XX  
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
PT applications including DNA immunization or generation of packaging cell  
PT lines, particularly in gene therapy.  
XX  
PS Example 1; Fig 6; 233pp; English.  
XX  
CC The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Env or Nef  
CC (1). (1) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunization, generation of packaging cell lines, and production of HIV  
CC type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABL40054 to ABL40215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 1509 BP; 321 A; 560 C; 470 G; 158 T; 0 U; 0 Other;

Query Match 98.8%; Score 1491.4; DB 6; Length 1509;  
Best Local Similarity 99.3%; Pred. No. 7.7e-175;  
Matches 1498; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 1 ATGGGCGCGCGCGAGCATCTGGCGCGGCAAGAGCTGACCAAGTGGAGAAATCCGC 60  
Db 1 ATGGGCGCGCGCGAGCATCTGGCGCGGCAAGAGCTGACCAAGTGGAGAAATCCGC 60  
QY 61 CTGGCGCGCGCGGCAAGAGCATCTGTAAGACCTGTGTGGGCGGCGGAG 120  
Db 61 CTGGCGCGCGCGGCAAGAGCATCTGTAAGACCTGTGTGGGCGGCGGAG 120  
QY 121 CTGGAGGCTTGGGCTTGAACCCCGGCTGTGAAGACCGCGGAGGCTGCAAGCATC 180  
Db 121 CTGGAGGCTTGGGCTTGAACCCCGGCTGTGAAGACCGCGGAGGCTGCAAGCATC 180  
QY 181 ATGAAGCATGTCAGCCCGCTTGAACCGGCAACCGAGAGCTGCGAGCTGTATCAAC 240  
Db 181 ATGAAGCATGTCAGCCCGCTTGAACCGGCAACCGAGAGCTGCGAGCTGTATCAAC 240  
QY 241 ACCGTGGCACTCTGTATCTGCTGTGAACCGGCAATCGAGTCCGCGACCAAGAGGCC 300  
Db 241 ACCGTGGCACTCTGTATCTGCTGTGAACCGGCAATCGAGTCCGCGACCAAGAGGCC 300  
QY 301 CTGGCAAGATCGAGGAGGAGCAAGCAAGTCCAGGAGAGAGCCAGGCGGCAAGAG 360  
Db 301 CTGGCAAGATCGAGGAGGAGCAAGCAAGTCCAGGAGAGAGCCAGGCGGCAAGAG 360  
QY 361 GCCGACGCGAGGTGAGCGAGCAACTATCTGTGAGAACTTGCAGGCGCGAGATGATG 420  
Db 361 GCCGACGCGAGGTGAGCGAGCAACTATCTGTGAGAACTTGCAGGCGCGAGATGATG 420  
QY 421 CACCAAGGCTATCAACCCCGGCAACCTTGAAGCGCTGTGAAGTGTATCGAGAGAGGCC 480  
Db 421 CACCAAGGCTATCAACCCCGGCAACCTTGAAGCGCTGTGAAGTGTATCGAGAGAGGCC 480  
QY 481 TTCAAGCCCGAGATGATCCCATGTTTCAACCGGCTGTGAAGGAGGCGCAACCCCGCA 540  
Db 481 TTCAAGCCCGAGATGATCCCATGTTTCAACCGGCTGTGAAGGAGGCGCAACCCCGCA 540  
QY 541 CTGAACAGATGTTGAACACCGTGGGCGGCCACCAAGGCGCGCATGACATGCTGAAGAC 600  
Db 541 CTGAACAGATGTTGAACACCGTGGGCGGCCACCAAGGCGCGCATGACATGCTGAAGAC 600  
QY 601 ACCATCAACGAGAGGCGCGCGAGTGGAGACCGCTTGAACCCCGTGAAGCGCGCGCTG 660  
Db 601 ACCATCAACGAGAGGCGCGCGAGTGGAGACCGCTTGAACCCCGTGAAGCGCGCGCTG 660  
QY 661 GCCCGCGCGAGATGCGCGACCCCGCGGAGCGAGCATTCGCGGCGCACCAAGACCTTG 720  
Db 661 GCCCGCGCGAGATGCGCGACCCCGCGGAGCGAGCATTCGCGGCGCACCAAGACCTTG 720  
QY 721 CAGGAGCAATCGCTGATGACCAAGCAACCCCGTGGCGGCGCATTTACAAAG 780  
Db 721 CAGGAGCAATCGCTGATGACCAAGCAACCCCGTGGCGGCGCATTTACAAAG 780  
QY 781 CCGTGATCATCTTGGGCGCTGAACAAGATGTAACAGTGAACAGCCCGGAGCATCTG 840  
Db 781 CCGTGATCATCTTGGGCGCTGAACAAGATGTAACAGTGAACAGCCCGGAGCATCTG 840  
QY 841 GACATTCGCGAGGCGCGGCAAGAGCTTTCGCGAGCTGAGTGAACGCTTTCAGAAC 900  
Db 841 GACATTCGCGAGGCGCGGCAAGAGCTTTCGCGAGCTGAGTGAACGCTTTCAGAAC 900  
QY 901 CTGGCGCGCGAGGCGGCAAGAGCTGTAAGAACTGATGACCGAGACCTTGTGTTG 960  
Db 901 CTGGCGCGCGAGGCGGCAAGAGCTGTAAGAACTGATGACCGAGACCTTGTGTTG 960  
QY 961 CAGAGCGGCAACCCCGAGCTGCAAGACCATCTGGCGGCTTGGGCGCGGCGGCAACCTG 1020  
Db 961 CAGAGCGGCAACCCCGAGCTGCAAGACCATCTGGCGGCTTGGGCGCGGCGGCAACCTG 1020  
QY 1021 GAGGAGATGATGACCGGCTGCGAGGCGTGGGCGGCGCGGCGCAAGGCGCGGCTG 1080



QY 781 CGGTGATCATCTCTGGGCTTGAACAAGATGTGCGATGTTACACCCCGTGAACATCTTG 840  
DB 781 CGCTGATCATCTCTGGGCTTGAACAAGATGTGCGATGTTACACCCCGTGAACATCTTG 840  
QY 841 GACATCGGCGGAGGCGCCCAAGAGAGCCCTTCCGCACTACGTGAGACCGCTTCTTCAAGAC 900  
DB 841 GACATCGGCGGAGGCGCCCAAGAGAGCCCTTCCGCACTACGTGAGACCGCTTCTTCAAGAC 900  
QY 901 CTGCGGCGGAGGCGCCCAAGAGAGCCCTTCCGCACTACGTGAGACCGCTTCTTCAAGAC 960  
DB 901 CTGCGGCGGAGGCGCCCAAGAGAGCCCTTCCGCACTACGTGAGACCGCTTCTTCAAGAC 960  
QY 961 CAGAGCGGCGGAGGCGCCCAAGAGAGCCCTTCCGCACTACGTGAGACCGCTTCTTCAAGAC 1020  
DB 961 CAGAGCGGCGGAGGCGCCCAAGAGAGCCCTTCCGCACTACGTGAGACCGCTTCTTCAAGAC 1020  
QY 1021 GAGGAGATGATGACCGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
DB 1021 GAGGAGATGATGACCGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
QY 1081 GCGGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
DB 1081 GCGGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
QY 1141 GCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
DB 1141 GCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
QY 1201 TGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260  
DB 1201 TGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260  
QY 1261 GACTGACCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
DB 1261 GACTGACCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
QY 1321 CCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
DB 1321 CCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
QY 1381 CCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
DB 1381 CCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
QY 1441 CCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
DB 1441 CCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
QY 1501 AGCCAGTAA 1509  
DB 1501 AGCCAGTAA 1509

RESULT 7  
AAAS1626  
ID AAAS1626 standard; DNA; 1509 BP.  
XX  
AC AAAS1626;  
XX  
DT 31-OCT-2000 (first entry)  
XX  
DE HIV codon-optimized synthetic Gag polynucleotide.  
XX  
XX Gag; expression cassette; antigenic; type C; HIV; Env; synthetic;  
XX DNA immunization; packaging cell line; antigen presentation; ss.  
OS Human immunodeficiency virus; type C strain AF110967.  
OS Synthetic.  
XX  
XX MO200039304-A2.  
XX  
XX 06-JUL-2000.  
PD

XX  
PF 30-DEC-1999; 99WO-US031273.  
PR 31-DEC-1998; 98US-0114495P.  
PR 01-SEP-1999; 99US-0152195P.  
XX  
PA (CHTR) CHIRON CORP.  
XX  
PI Barnett S, Zur Megede J;  
XX  
DR WPI; 2000-452401/39.  
XX  
PT Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env  
PT polypeptide and the polypeptide useful for immunizing a mammal especially  
PT human against HIV.  
XX  
PS Disclosure; Page 104; 113pp; English.  
XX  
CC Expression cassette comprising a polynucleotide encoding antigenic type  
CC C human immunodeficiency virus (HIV) Gag or Env polypeptides are useful  
CC in DNA immunization, generation of packaging cell lines and production of  
CC Gag- and/or Env-containing proteins. Synthetic Env and Gag expression  
CC cassettes exhibit increased potency for induction of cytotoxic T-  
CC lymphocyte (CTL) responses by DNA immunization. Gag of HIV-1 self-  
CC assemble into non-infectious virus-like particles which are used as a  
CC matrix for the proper presentation of an antigen entrapped or associated  
CC to the immune system of the host  
XX  
SQ Sequence 1509 BP; 321 A; 559 C; 471 G; 158 T; 0 U; 0 Other;  
XX  
Query Match 98.7%; Score 1489.8; DB 3; Length 1509;  
Best Local Similarity 99.2%; Pred. No. 1.2e-174;  
Matches 1497; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 1 ATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60  
DB 1 ATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60  
QY 61 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
DB 61 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
QY 121 CTGGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180  
DB 121 CTGGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180  
QY 181 ATGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
DB 181 ATGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
QY 241 ACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
DB 241 ACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
QY 301 CTGGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
DB 301 CTGGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
QY 361 GCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
DB 361 GCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
QY 421 CACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
DB 421 CACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
QY 481 TTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540  
DB 481 TTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540  
QY 541 CTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
DB 541 CTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600



Sequence 1509 BP; 321 A; 559 C; 471 G; 158 T; 0 U; 0 Other;

Query Match 98.7%; Score 1489.8; DB 6; Length 1509;  
Best Local Similarity 99.2%; Pred. No. 1.2e-174;  
Matches 1497; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
QY 1 ATGGGGGCGCGCGCGAGCATCTCGCGCGCGAGAGCAAGTGGGAGGAATCCCG 60
DB 1 ATGGGGGCGCGCGCGAGCATCTCGCGCGCGAGAGCAAGTGGGAGGAATCCCG 60
QY 61 CTGCGCGCGCGCGCGCAAGAGCACTACATGCTGAACACCTGGTGTGGCCAGCCGCGAG 120
DB 61 CTGCGCGCGCGCGCGCAAGAGCACTACATGCTGAACACCTGGTGTGGCCAGCCGCGAG 120
QY 121 CTGGAAGGGCTTGCGCCTGTGAACCCCGGCTGTGTGAAGACCGCGAGGGCTTGCAAGCATC 180
DB 121 CTGGAAGGGCTTGCGCCTGTGAACCCCGGCTGTGTGAAGACCGCGAGGGCTTGCAAGCATC 180
QY 181 ATGAAGCAGCTGAGCGCGCGCTGAGACCGGACCGAGAGAGTGGCGAGCCTGTACAC 240
DB 181 ATGAAGCAGCTGAGCGCGCGCTGTGACCGGACCGAGAGAGTGGCGAGCCTGTACAC 240
QY 241 ACCGTGGCCACCTGTACTGTGTGACGCGCGCATGAGGTCCGCGACCAAGAGAGCC 300
DB 241 ACCGTGGCCACCTGTACTGTGTGACGCGCGCATGAGGTCCGCGACCAAGAGAGCC 300
QY 301 CTGGAACAAATGAGAGAGAGAGCAAGAACTCCACAGAAAGCCAGAGGCCAAGAG 360
DB 301 CTGGAACAAATGAGAGAGAGAGCAAGAACTCCACAGAAAGCCAGAGGCCAAGAG 360
QY 361 GCGGAGCGGAGAGGTGAGCCAGAACTACCCCATGTGACAGACTGAGGGCGCAGATGGT 420
DB 361 GCGGAGCGGAGAGGTGAGCCAGAACTACCCCATGTGACAGACTGAGGGCGCAGATGGT 420
QY 421 CACAGAGCCATCAGCGCGCGCGACCTGTGAAGCGCTGGGTGAAGTATCGAGAGAGAGCC 480
DB 421 CACAGAGCCATCAGCGCGCGCGACCTGTGAAGCGCTGGGTGAAGTATCGAGAGAGAGCC 480
QY 481 TTCAAGCGCGGAGGTATCTCCCATGTTCACCGCGCTGAGAGAGGGCGCCACCCCGCAGAC 540
DB 481 TTCAAGCGCGGAGGTATCTCCCATGTTCACCGCGCTGAGAGAGGGCGCCACCCCGCAGAC 540
QY 541 CTGAACACGATGTGAAACACCGTGGGCGGCGACAGGCGCGCATGAGATGCTGAAGAGC 600
DB 541 CTGAACACGATGTGAAACACCGTGGGCGGCGACAGGCGCGCATGAGATGCTGAAGAGC 600
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCGGCTGTGACCCCGTGCAGGGCCGCGCTG 660
DB 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCGGCTGTGACCCCGTGCAGGGCCGCGCTG 660
QY 661 GCGCGCGCGGAGATGGCGGAGACCCCGCGCGAGCATGCGGCGCGCACCGACCCCTG 720
DB 661 GCGCGCGCGGAGATGGCGGAGACCCCGCGCGAGCATGCGGCGCGCACCGACCCCTG 720
QY 721 CAGGAGCAGATGCGCTGATGACAGCAACCCCGCGTGCAGGGCGCGCATCTACAAG 780
DB 721 CAGGAGCAGATGCGCTGATGACAGCAACCCCGCGTGCAGGGCGCGCATCTACAAG 780
QY 781 CGGTGATCATCTTGGGCTTGAACAAAGATGTCGAGATGTACAGCCCGGTGAGCATCTTG 840
DB 781 CGGTGATCATCTTGGGCTTGAACAAAGATGTCGAGATGTACAGCCCGGTGAGCATCTTG 840
QY 841 GACATCCGCGAGGGGCGCGAAGAGCCCTTCGCGCATCTGTCGAGACCGCTTCTTCAAGAC 900
DB 841 GACATCCGCGAGGGGCGCGAAGAGCCCTTCGCGCATCTGTCGAGACCGCTTCTTCAAGAC 900
QY 901 CTGCGCGCGGAGCAGAGCCAGAGAGCTGAAGAACTGGATGACCGAGACCTGTGTGGT 960
DB 901 CTGCGCGCGGAGCAGAGCCAGAGAGCTGAAGAACTGGATGACCGAGACCTGTGTGGT 960
QY 961 CAGAAAGCCAGACCCGAGCTGCAAGACCATCTGCGCGCTTGCAGCCCGCGCGCACCTG 1020
DB 961 CAGAAAGCCAGACCCGAGCTGCAAGACCATCTGCGCGCTTGCAGCCCGCGCGCACCTG 1020
```

```
QY 1021 GAGGATGATGATGACCGCTGCGAGGGCGTGGCGGCGCGCGCGCACAGGCCCGCTGTG 1080
DB 1021 GAGGATGATGATGACCGCTGCGAGGGCGTGGCGGCGCGCGCGCACAGGCCCGCTGTG 1080
QY 1081 GCGGAGGGGATGAGCAGGCGCAACAGCGTGAACATCATGATGCAAGAGCACTTCAAG 1140
DB 1081 GCGGAGGGGATGAGCAGGCGCAACAGCGTGAACATCATGATGCAAGAGCACTTCAAG 1140
QY 1141 GCGCGCGCGCGCAACGTCATGCTTCAACTGCGCAAGAGGGCCATCGCCCAAGAAC 1200
DB 1141 GCGCGCGCGCGCAACGTCATGCTTCAACTGCGCAAGAGGGCCATCGCCCAAGAAC 1200
QY 1201 TGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGTGCGCAAGAGAGGCCACCAAGTGAAG 1260
DB 1201 TGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGTGCGCAAGAGAGGCCACCAAGTGAAG 1260
QY 1261 GACTGCAACGAGCGCGCAAGGCCAATCTTGGGCAAGATCTGGCCCAAGAGGCGCG 1320
DB 1261 GACTGCAACGAGCGCGCAAGGCCAATCTTGGGCAAGATCTGGCCCAAGAGGCGCG 1320
QY 1321 CCGGCGCACTTCTGCAAGACCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
DB 1321 CCGGCGCACTTCTGCAAGACCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
QY 1381 CCGCGCGAGAGCTTCTGCGCTTGAAGAGACACCCCGCGCGCGCGCGCGCGCGCGCG 1440
DB 1381 CCGCGCGAGAGCTTCTGCGCTTGAAGAGACACCCCGCGCGCGCGCGCGCGCGCGCG 1440
QY 1441 GCGGAGCGCTTACCGGAGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
DB 1441 GCGGAGCGCTTACCGGAGCGCGCTTGAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
QY 1501 AGCCAGTAA 1509
DB 1501 AGCCAGTAA 1509
RESULT 9
ABL39972
ID ABL39972 standard; DNA; 1494 BP.
XX
XX ABL39972;
AC
XX
XX 15-MAY-2002 (first entry)
DT
XX
DE Synthetic Gag polynucleotide sequence SEQ ID NO:51.
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; env; nef;
KW immunostimulant; gene therapy; gene; ds.
OS
OS Human immunodeficiency virus; type C.
XX
XX Synthetic.
XX
XX MO200204493-A2.
XX
XX 17-JAN-2002.
PD
XX
XX 05-JUL-2001; 2001WO-US021241.
PF
XX
XX 05-JUL-2000; 2000US-00610313.
PR
XX
XX (CHIR ) CHIRON CORP.
PA
XX (UYST-) UNIV STELLENBOSCH.
XX
XX Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
PI
XX WPI; 2002-154920/20.
XX
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
```

XX Claim 7; Fig 22; 233bp; English.  
PS  
CC The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (i). (i) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL3942 to ABL40054 and  
CC ABL06204 to ABL06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 1494 BP; 325 A; 557 C; 455 G; 157 T; 0 U; 0 Other;  
  
Query Match 86.0%; Score 1297.6; DB 6; Length 1494;  
Best Local Similarity 92.7%; Pred. No. 4,7e-151;  
Matches 1402; Conservative 0; Mismatches 89; Indels 21; Gaps 3;  
  
QY 1 ATGGGCGCCCGCCGAGCATCTGCGCGGCGAGAGCTGAGCAAGTGGAGAGATCCGC 60  
DB 1 ATGGGCGCCCGCCGAGCATCTGAGCGGCGGCACTGAGCAAGTGGAGCGCATCCGC 60  
QY 61 CTGCGCGCCCGGCGGCAAGAGCACTAATGCTGAGCACTGCTGTGGCCAGCCGCGAG 120  
DB 61 CTGCGCGCCCGGCGGCAAGAGCACTAATGCTGAGCACTGCTGTGGCCAGCCGCGAG 120  
QY 121 CTGGAAGGCTTGCGCCCTGAAACCCCGGCTGTGAGAGACCGCGAGGGCTGCAAGCAGATC 180  
DB 121 CTGGAAGGCTTGCGCCCTGAAACCCCGGCTGTGAGAGACGAGGAGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTCAGCCCGCCCTGCGAGACCGGACCGAGAGAGCTGCGACCTGTATCAAC 240  
DB 181 ATCAAGCAGCTCAGCCCGCCCTGCGAGACCGGACCGAGAGAGCTGCGACCTGTATCAAC 240  
QY 241 ACCGTGGCCACCTCTGTATCTGCTGTGACGCGCGGATGAGAGTCCGGACACCAAGAGGCC 300  
DB 241 ACCGTGGCCACCTCTGTATCTGCTGTGACGAGGGATGAGAGTGGCGACCAAGAGGCC 300  
QY 301 CTGGAACAAGATCGAGGAGAGCAAGCAAGTCCACAGCAAGAGCCAGAGAGCCAGAGAG 360  
DB 301 CTGGAACAAGATCGAGGAGAGCAAGCAAGTCCACAGCAAGAGCCAGAGAGCCAGAGAG 360  
QY 361 GCGGACGGAAGTGAAGCACTAATCCCATGTGCAAGACCTGCAAGGCGCAAGATGCTG 420  
DB 361 GCGGACGGAAGTGAAGCACTAATCCCATGTGCAAGACGCGGCGCAAGATGCTG 420  
QY 421 CACCAAGGCGCATCAGCCCGGCAACCTGAAAGCGCTGTATCAAGGTATCAAGAGAGGCC 480  
DB 421 CACCAAGGCGCATCAGCCCGGCAACCTGAAAGCGCTGTATCAAGGTATCAAGAGAGGCC 480  
QY 481 TTCAGCGCCGAGAGTATCCCATGTTCACGCGCTGAGGAGAGGGCGCAACCCCGCAGGAC 540  
DB 481 TTCAGCGCCGAGAGATCCCATGTTCACGCGCTGAGGAGAGGGCGCAACCCCGCAGGAC 540  
QY 541 CTGAACACGATGTGAACAACCGTGGGCGGCAACAGGCCGCGCATGTGCTGAAGAGAC 600  
DB 541 CTGAACACGATGTGAACAACCGTGGGCGGCAACAGGCCGCGCATGTGCTGAAGAGAC 600  
QY 601 ACCATCAACGAGAGAGCGCCCGGAGTGGAGCCGCTGACACCCCGTGTGAGGCCGCGCTG 660  
DB 601 ACCATCAACGAGAGAGCGCCCGGAGTGGAGCCGCTGACACCCCGTGTGAGGCCGCGCTG 660  
QY 661 GCGCGCGGCGAGATGGCGGAGCCCGGCGGAGGAGATGCGGCGGCAACCGGCGCTG 720  
DB 661 GCGCGCGGCGAGATGGCGGAGCCCGGCGGAGGAGATGCGGCGGCAACCGGCGCTG 720  
QY 721 CAGAGCAGATCGCTGATGAACAGCAACCCCGCGTGCCTGTGGGCGACATCTTAAG 780  
DB 721 CAGAGCAGATCGCTGATGAACAGCAACCCCGCGTGCCTGTGGGCGACATCTTAAG 780

QY 781 CGGTGATATCTTGGGCGCTGAACAAGATCGGAGATGTATAGCCCGGTGAGCATCCG 840  
DB 781 CGGTGATATCTTGGGCGCTGAACAAGATCGGAGATGTATAGCCCGGTGAGCATCCG 840  
QY 841 GACATCCGCAAGGCGCCCAAGAGGCGCTTCCGCACTACGTGAGCGCTTCTTCAAGACC 900  
DB 841 GACATCAAGAGGCGCCCAAGAGGCGCTTCCGCACTACGTGAGCGCTTCTTCAAGACC 900  
QY 901 CTGCGCGCGGAGAGGCGCAACCGAGACGTTGAAGAACTGGATATGACGAGACCTGCTGTG 960  
DB 901 CTGCGCGCGGAGAGGCGCAACCGAGACGTTGAAGAACTGGATATGACGAGACCTGCTGTG 960  
QY 961 CAGAACGCGCAACCCGAGCTGAGAGCAATCCGCGCGCTGTGGCCCGGCGCAACCCG 1020  
DB 961 CAGAACGCGCAACCCGAGCTGAGAGCAATCCGCGCGCTGTGGCCCGGCGCAACCCG 1020  
QY 1021 GAGAGATGATGACCGCTGCGAGGCGGTGGGCGGCGCCGCGCAGAGGCGCGGTGCTG 1080  
DB 1021 GAGAGATGATGACCGCTGCGAGGCGGTGGGCGGCGCCGCGCAGAGGCGCGGTGCTG 1080  
QY 1081 GCGGAGCGGATGAGCGCAAGCGCAACGCTGAAATCATGATGATGACAGAGCACTTCAAG 1140  
DB 1081 GCGGAGCGGATGAGCGCAAGCGCAACGCTGAAATCATGATGATGACAGAGCACTTCAAG 1137  
QY 1141 GCGCGCGCGGCGCAAGCTGAGTCTTCAATCGCGCGCAAGAGGCGCAATGCGCAAGAAC 1200  
DB 1138 GCGAGCAACCGCATATCAAGTCTTCAATCGCGCGCAAGAGTGGGCGCAATGCGCGCAAC 1197  
QY 1201 TGCGCGCGCGCGCAAGAGGCGCTGAGAAAGTGGCGCAAGAGAGGCGCAAGATGAAG 1260  
DB 1198 TGCGCGCGCGCGCAAGAGGCGCTGAGAAAGTGGCGCAAGAGGCGCAAGATGAAG 1257  
QY 1261 GACTCACCAGAGCGCGCAAGCTTCTGTGGCAAGATCTGGCCAGCCAGCAAGAGGCGCG 1320  
DB 1258 GACTCACCAGAGCGCGCAAGCTTCTGTGGCAAGATCTGGCCAGCCAGCAAGAGGCGCG 1317  
QY 1321 CCGGCGCACTTCTGTGAAGACCGGCGAGCGAGC---CGGCGCGCGCGCAAGTGGCGCGCG 1377  
DB 1318 CCGGCGCACTTCTGTGAAGACCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377  
QY 1378 CCGCGCGCGGAGGCTTCCGCTTGAAGAGACCAACCGCGCGCGCGCGCGCGCGCGCG 1437  
DB 1378 CCGCGCGCGGAGGCTTCCGCTTGAAGAGACCAACCGCGCGCGCGCGCGCGCGCGCG 1437  
QY 1438 GACGCGAGCGCTTACCGGAGCGCTTGAACCGCGCTTGGCGAGCTTGGCGAGCGCGCG 1497  
DB 1438 -----CGGAGCGCGCTGACCAAGCTGAGAGAGCGCTTGGCGAGCGAGCGCG 1482  
QY 1498 CTGAGCCAGTAA 1509  
DB 1483 CTGAGCCAGTAA 1494  
  
RESULT 10  
ADM73785  
ID ADM73785 standard; DNA, 1494 BP.  
XX  
XX ADM73785;  
XX  
XX 03-JUN-2004 (first entry)  
XX  
XX  
XX HIV-1 polynucleotide #20.  
XX  
XX HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
XX  
XX HIV type C protein; immunostimulant.  
XX  
XX Human immunodeficiency virus 1.  
XX  
XX US200323961-A1.  
XX  
XX 04-DEC-2003.  
XX  
XX 05-JUL-2001; 2001US-0089575.  
PF



XX 05-JUL-2000; 2000US-00610313.  
 PR (MEGE/) MEGEDE J Z.  
 XX (BARN/) BARNETT S W.  
 PA (ENGE/) ENGELBRECHT S.  
 PA (RENS/) RENSBERG E J V.  
 XX  
 PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV,  
 DR WPI; 2004-060515/06.  
 XX  
 PT New expression cassette comprising a polynucleotide sequence encoding an  
 PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
 PT immunization, generating of packaging cell lines or in producing HIV Type  
 PT C proteins.  
 XX  
 PS Claim 7; SEQ ID NO 51, 160bp; English.  
 XX  
 CC The invention relates to an expression cassette comprising a  
 CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention  
 CC also relates to a recombinant expression system for use in a host cell  
 CC comprising an expression cassette, where the polynucleotide sequence  
 CC further comprises control elements capable of driving expression in the  
 CC selected host cell, a cell comprising an expression cassette where the  
 CC polynucleotide sequence further comprises control elements compatible  
 CC with the expression in the cell and a composition for generating an  
 CC immunological response, comprising an expression cassette. The expression  
 CC cassette and the methods of the invention are useful in eliciting an  
 CC immune response, in DNA immunisation, in generation of packaging cell  
 CC lines and in producing HIV Type C proteins. This sequence represents an  
 CC HIV-1 polynucleotide of the invention.  
 XX  
 SO Sequence 1494 BP; 325 A; 557 C; 455 G; 157 T; 0 U; 0 Other;  
 Query Match 86.0%; Score 1297.6; DB 12; Length 1494;  
 Best Local Similarity 92.7%; Pred. No. 4.7e-151;  
 Matches 1402; Conservative 0; Mismatches 89; Indels 21; Gaps 3;

Db 481 TTCAAACCCCGAGAGATCCCATGTTCAACCGCTTGAACGAGGCGCAACCCCGAGAGC 540  
 Qy 541 CTGAACAGATGTTGAACACCGTGGCGGCAACGAGCGGCATGATGTTGAAGAC 600  
 Db 541 CTGAACAGATGTTGAACACCGTGGCGGCAACGAGCGGCATGATGTTGAAGAC 600  
 Qy 601 ACCATCAACAGAGAGCGCGGAGTGGAGACCGCTGCAACCCCGTGAAGAGCGCGCGCTG 660  
 Db 601 ACCATCAACAGAGAGCGCGGAGTGGAGACCGCAACCCCGTGAAGAGCGCGCGCGCTG 660  
 Qy 661 GCCCGGCGAGATGCGGAGACCCCGGCGGAGCAATCGCGGCGGCAACGAGACCTG 720  
 Db 661 GCCCGGCGAGATGCGGAGACCCCGGCGGAGCAATCGCGGCGGCAACGAGACCTG 720  
 Qy 721 CAGAGCAGATGCGCTGATGACAGCAACCCCGGCGGAGTGGAGCAATGATGACAG 780  
 Db 721 CAGAGCAGATGCGCTGATGACAGCAACCCCGGCGGAGTGGAGCAATGATGACAG 780  
 Qy 781 CGTGGATCATCTGGGCTTGAACAGATCGTGGATGTAACAGCCCGTGAAGATCCTG 840  
 Db 781 CGTGGATCATCTGGGCTTGAACAGATCGTGGATGTAACAGCCCGTGAAGATCCTG 840  
 Qy 841 GACATCCGCGAGAGCGCGGAGAGCGCTTGGAGCACTACGAGACCGCTTCTTAAGAC 900  
 Db 841 GACATCCGCGAGAGCGCGGAGAGCGCTTGGAGCACTACGAGACCGCTTCTTAAGAC 900  
 Qy 901 CTGGCGCGGAGAGCGCGGAGAGCGCTTGGAGCACTACGAGACCGCTTCTTAAGAC 960  
 Db 901 CTGGCGCGGAGAGCGCGGAGAGCGCTTGGAGCACTACGAGACCGCTTCTTAAGAC 960  
 Qy 961 CAGAACGCAACCCCGAGTGAAGCAATCGTGGCGCTTGGAGCACTACGAGACCGCTT 1020  
 Db 961 CAGAACGCAACCCCGAGTGAAGCAATCGTGGCGCTTGGAGCACTACGAGACCGCTT 1020  
 Qy 1021 GAGAGATGATGACCGGCTTGAACAGATCGTGGATGTAACAGCCCGTGAAGATCCTG 1080  
 Db 1021 GAGAGATGATGACCGGCTTGAACAGATCGTGGATGTAACAGCCCGTGAAGATCCTG 1080  
 Qy 1081 GCCAGGCGATGAGCGCGGAGAGCGCTTGAACAGATCGTGGATGTAACAGCCCGT 1140  
 Db 1081 GCCAGGCGATGAGCGCGGAGAGCGCTTGAACAGATCGTGGATGTAACAGCCCGT 1140  
 Qy 1141 GCGCGCGCGGAGAGCGCGGAGAGCGCTTGAACAGATCGTGGATGTAACAGCCCGT 1200  
 Db 1138 GCGAGCAACCGCATCATGATGCTTCACTGGCGGCAAGGTGGGCCCATGCGCCGCAAC 1197  
 Qy 1201 TGCGCGCGCGGAGAGAGCGCTTGAACAGATCGTGGATGTAACAGCCCGTGAAG 1260  
 Db 1198 TGCGCGCGCGGAGAGAGCGCTTGAACAGATCGTGGATGTAACAGCCCGTGAAG 1257  
 Qy 1261 GACTGCAACGAGGCGCGGAGAGCGCTTGAACAGATCGTGGATGTAACAGCCCGT 1320  
 Db 1258 GACTGCAACGAGGCGCGGAGAGCGCTTGAACAGATCGTGGATGTAACAGCCCGT 1317  
 Qy 1321 CCGGCGCACTTCTGCAAGACCGAGAGCG---CGCGCGCGCGCAACCGGCGCAACCG 1377  
 Db 1318 CCGGCGCACTTCTGCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377  
 Qy 1378 CCGCGCGCGGAGAGCTTCTGCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437  
 Db 1378 CCGCGCGCGGAGAGCTTCTGCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437  
 Qy 1438 GACCGGAGCGCTTCTGCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1497  
 Db 1438 -----GCGAGCGCGCTGACCAAGCTGTAAGAGCTGTTGGCAAGCGACCGC 1482  
 Qy 1498 CTGAGCCAGTAA 1509  
 Db 1483 CTGAGCCAGTAA 1494  
 RESULT 11  
 ABL40020



ID ABL40020 standard; DNA; 1491 BP.  
XX ABL40020;  
AC  
XX  
DT 15-MAY-2002 (first entry)  
XX  
XX  
DE Synthetic Gag polynucleotide sequence SEQ ID NO:99.  
XX  
XX  
KW Human immunodeficiency virus type C; antigenic HIV type C protein;  
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
KW immunostimulant; gene therapy; gene; ds.  
XX  
OS Human immunodeficiency virus; type C.  
OS Synthetic.  
XX  
FN WO200204493-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 05-JUL-2001; 2001WO-US021241.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megele J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
DR WPI, 2002-154920/20.  
XX  
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
PT applications including DNA immunisation or generation of packaging cell  
PI lines, particularly in gene therapy.  
XX  
XX  
PS Claim 8; Fig 70; 233pp; English.  
XX  
XX  
CC The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (1). (i) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL3942 to ABL4004 and  
CC ABL06204 to ABL06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 1491 BP; 319 A; 557 C; 457 G; 158 T; 0 U; 0 Other;  
Query Match 85.8%; Score 1294.8; DB 6; Length 1491;  
Best Local Similarity 92.7%; Pred. No. 1e-150;  
Matches 1401; Conservative 0; Mismatches 87; Indels 24; Gaps 3;  
QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGCGGAGAGTGGAGGAAGATCCGC 60  
DB 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGCGGAGTGGAGGAAGATCCGC 60  
QY 61 CTGCGCGCGCGCGCGCAAGAGCATCAATGCTGAAGCACTGTGTGGCGCGCGAG 120  
DB 61 CTGCGCGCGCGCGCGCAAGAGCATCAATGCTGAAGCACTGTGTGGCGCGCGAG 120  
QY 121 CTGGAAGGCTTGTGCTGAAACCCCGGCTGTGGAAGACCGCGAGGGCTGCAAGATC 180  
DB 121 CTGGAAGGCTTGTGCTGAAACCCCGGCTGTGGAAGACCGCGAGGGCTGCAAGATC 180  
QY 121 CTGGAAGGCTTGTGCTGAAACCCCGGCTGTGGAAGACCGCGAGGGCTGCAAGATC 180  
DB 121 CTGGAAGGCTTGTGCTGAAACCCCGGCTGTGGAAGACCGCGAGGGCTGCAAGATC 180  
QY 181 ATGAAGCAGCTGCAAGCGCGCGCTGCAAGACCGGAGAGCTGGGAGCTGTAAAC 240  
DB 181 ATGAAGCAGCTGCAAGCGCGCGCTGCAAGACCGGAGAGCTGGGAGCTGTAAAC 240  
QY 181 ATCAAGCAGCTGCAAGCGCGCGCTGCAAGACCGGAGAGATCGGAGCTGTAAAC 240  
DB 181 ATCAAGCAGCTGCAAGCGCGCGCTGCAAGACCGGAGAGATCGGAGCTGTAAAC 240  
QY 241 ACCGTGGCACCCTTACTGCTGTCGTCGACCGCGCATGAGTCCGGACACCAAGAGGCC 300  
DB 241 ACCGTGGCACCCTTACTGCTGTCGTCGACCGCGCATGAGTCCGGACACCAAGAGGCC 300

QY 301 CTGGAAGATCGAAGAGAGAGAGAAACAAGTCCCAAGAGAAACCCAGAGGCCAAGAG 360  
DB 301 CTGGAAGATCGAAGAGAGAGAGAAACAAGTCCCAAGAGAAACCCAGAGGCCAAGAG 360  
QY 361 GCCGACGGCAAGGTGAGCCAGAACTAACCCATCTGTGCAAGAACTTCAGGGCCAGATG 420  
DB 361 GCCGACGAAGAGGTGAGCCAGAACTAACCCATCTGTGCAAGAACTTCAGGGCCAGATG 420  
QY 421 CACGAGGCTATGAGCCCCCGCAACCTGAAACGCTGTGGTGAAGGTATCGAAGAGAGGCC 480  
DB 421 CACGAGGCTATGAGCCCCCGCAACCTGAAACGCTGTGGTGAAGGTATCGAAGAGAGGCC 480  
QY 481 TTGAGCCCCGAGAGTATCCCAATGTTCAACCGCCCTGAGACGAGGGGCCACCCCAAGAC 540  
DB 481 TTGAGCCCCGAGAGTATCCCAATGTTCAACCGCCCTGAGACGAGGGGCCACCCCAAGAC 540  
QY 541 CTGAACACGATGTTGAACAACGCTGGGCGGCAACAGCGCGCCATGCAAGTCTGAAGAC 600  
DB 541 CTGAACACGATGTTGAACAACGCTGGGCGGCAACAGCGCGCCATGCAAGTCTGAAGAC 600  
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGACCGCTGCAACCCCGTCAGCGCGGCCCGTG 660  
DB 601 ACCATCAACGAGAGAGCGCGCGAGTGGACCGCTGCAACCCCGTCAGCGCGGCCCGTG 660  
QY 661 GCCCGCGCGCAATGCGCGGACCCCGCGCGGCAATCGCGCGGCCACCAAGCAACCTTG 720  
DB 661 GCCCGCGCGCAATGCGCGGACCCCGCGCGGCAATCGCGCGGCCACCAAGCAACCTTG 720  
QY 721 CAGAGCAGATCGCTGATGACAGCAACCCCGCGCGCGTGGCGGCACTTCAAG 780  
DB 721 CAGAGCAGATCGCTGATGACAGCAACCCCGCGCGCGTGGCGGCACTTCAAG 780  
QY 781 CGGTGATCATCTGTGGCTTGAACAAGATCGTGCAGATGTAACGCCCGTGAAGATCTTG 840  
DB 781 CGGTGATCATCTGTGGCTTGAACAAGATCGTGCAGATGTAACGCCCGTGAAGATCTTG 840  
QY 841 GACATCCGCGAGAGGCCCAAGAGACCTTCCGCGCACTACGTGACCGCTTCTTCAAGAC 900  
DB 841 GACATCCGCGAGAGGCCCAAGAGACCTTCCGCGCACTACGTGACCGCTTCTTCAAGAC 900  
QY 901 CTGCGCGCGCGAGAGGCCACCAAGAGCTGGAAGAACTGATGACCAACCTTGTG 960  
DB 901 CTGCGCGCGCGAGAGGCCACCAAGAGCTGGAAGAACTGATGACCAACCTTGTG 960  
QY 961 CAGAACGCAACCCCGCATGCAAGACATCTGCGCGCTTCCGCGCGCGCGCAACCTTG 1020  
DB 961 CAGAACGCAACCCCGCATGCAAGACATCTGCGCGCTTCCGCGCGCGCGCAACCTTG 1020  
QY 1021 GAGGATGATGATCGCTGCGCAAGGCGGTGGCGCGGCCCGGCAAGGCCCGGTGCTG 1080  
DB 1021 GAGGATGATGATCGCTGCGCAAGGCGGTGGCGCGGCCCGGCAAGGCCCGGTGCTG 1080  
QY 1081 GCCGAGGAGATGAGCGCAGGCCCAACAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140  
DB 1081 GCCGAGGAGATGAGCGCAGGCCCAACAGCGTGAACATCATGATGACAGAGCAACTTCAAG 1140  
QY 1141 GGGCCCCCGCGCAACGTCAGTCTTCAACTGCGGCAAGAGAGGCCACATGCGCAAGAC 1200  
DB 1141 GGGCCCCCGCGCGCTGTAAGTCTTCAACTGCGGCGCGGAGGCCACATGCGCGCAAC 1200  
QY 1201 TGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGTGCGCAAGAGAGGCCACATGTAAG 1260  
DB 1201 TGCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGTGCGCAAGAGAGGCCACATGTAAG 1260  
QY 1261 GACTGCAACGAGCGCGCAAGCACTTCTGTGGCAAGATCTGAGCCAGCAACAGGGCGCG 1320  
DB 1261 GACTGCAACGAGCGCGCGCAAGCACTTCTGTGGCAAGATCTGAGCCAGCAACAGGGCGCG 1320  
QY 1321 CCGGCAACTTCTGTGCAAGACCGAGCGAGCGCGCGCGCGCACG--GTGCCACGCGC 1377  
DB 1321 CCGGCAACTTCTGTGCAAGACCGCGAGCGCGCGCGCGCGCGCACGCGCGC 1380

```

QY 1378 CCCCCCGCGAGAGCTTCGCGCTTCGAGAGACCAACCCCCCGCCAGACGAGAGCCCAAG 1437
Db 1381 CCCCCCGCGAGAGCTTCAGAGTTCAAGAGAGA-----CCCCCAAGCAGAGCCCAAG 1431
QY 1438 GACCGGAGAGCCCTACCGGAGAGCCCTGACCGCCCTGCGCAGAGCTGTTGGAGAGGGCCCC 1497
Db 1432 G-----ACCGGAGAGCCCTGACAGCCTGTAAGAGAGCTGTTGGAGAGGAGCCCC 1479
QY 1498 CTGAGCCCACTAA 1509
Db 1480 CTGAGCCCACTAA 1491

RESULT 12
ADM73833
ID ADM73833 standard; DNA; 1491 BP.
XX ADM73833;
XX AC
XX AD
XX 03-JUN-2004 (first entry)
XX DE
XX HIV-1 polynucleotide #68.
XX KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
XX KW HIV type C protein; immunostimulant.
XX OS
XX Human immunodeficiency virus 1.
XX PN US2003223961-A1.
XX PD
XX 04-DEC-2003.
XX PF 05-JUL-2001; 2001US-00899575.
XX PR 05-JUL-2000; 2000US-00610313.
XX PA (MEGE/) MEGEDE J Z.
XX PA (BARN/) BARNETT S W.
XX PA (ENGE/) ENGELBRECHT S.
XX PA (RENS/) RENSBURG E J V.
XX PI
XX PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX DR WPI: 2004-060515/06.
XX
XX New expression cassette comprising a polynucleotide sequence encoding an
XX HIV Pol polypeptide, useful in eliciting an immune response, in DNA
XX immunisation, generating of packaging cell lines or in producing HIV type
XX C proteins.
XX
XX Claim 8; SEQ ID NO 99; 160pp; English.
XX
XX The invention relates to an expression cassette comprising a
XX polynucleotide sequence encoding an HIV Pol polypeptide. The invention
XX also relates to a recombinant expression system for use in a host cell
XX comprising an expression cassette, where the polynucleotide sequence
XX further comprises control elements capable of driving expression in the
XX selected host cell, a cell comprising an expression cassette where the
XX polynucleotide sequence further comprises control elements compatible
XX with the expression in the cell and a composition for generating an
XX immunological response, comprising an expression cassette. The expression
XX cassette and the methods of the invention are useful in eliciting an
XX immune response, in DNA immunisation, in generation of packaging cell
XX lines and in producing HIV type C proteins. This sequence represents an
XX HIV-1 polynucleotide of the invention.
XX
XX Sequence 1491 BP; 319 A; 557 C; 457 G; 158 T; 0 U; 0 Other;
XX
XX Query Match 85.8%; Score 1294.8; DB 12; Length 1491;
XX Best Local Similarity 92.7%; Pred. No.1e-150;
XX Matches 1401; Conservative 0; Mismatches 87; Indels 24; Gaps 3;
XX
XX 1 ATGGGCGCCCGCGAGCATCTGCGCGGAGAGCTGAGCAAGTGGAGAGATCCGC 60

```

Db	1	ATGGGGGCGCCGCGCCAGCAATCTCGCGCGCGCAACTGGACAAAGTGGAGAAAGATCCGG	60
Qy	61	CTGGGCGCGCGCGCGCAAGAAACACTACATCTGAACAACACTGGTGTGGCCAGCGCGGAG	120
Db	61	CTGGCGCCCGCGCGCGCGCAAGCACTACATCTGAACAACACTGGTGTGGCCAGCGCGGAG	120
Qy	121	CTGGAGGAGCTTCGCTCTGAACTCCCGGCTTGTGAAGCCGCGAGGCGTGGCAAGCATC	180
Db	121	CTGAGCGCTTCGCTCTGAACTCCCGGCTTGTGAAGCCGCGAGGCGTGGCAAGCATC	180
Qy	181	ATGAACCACTGAAGCGCCCTCTGGACACCGGCACTGAAGAGTGTGGACCTGTACAAAC	240
Db	181	ATCAACCACTGTGAAGCGCCCTCTGGACACCGGCACTGAAGAGTGTGGACCTGTTCAAAC	240
Qy	241	ACCGTGGCCACCTGTACTGCTGTCAGCGCGGCACTGAAGTCCGCGCAACAAGAGAGCC	300
Db	241	ACCGTGGCCACCTGTACTGCTGTCAGCAAGGCGCATGACGTGCGGCAACAAGAGAGCC	300
Qy	301	CTGGACACAGATGAGAGAGAGCAAAAGTCCACAGAGAGACCCACAGGCGCAAGAG	360
Db	301	CTGGACACAGATGAGAGAGAGCAAAAGTCCACAGAGAGACCCACAGAGCGCGAGGCC	360
Qy	361	GGCGACGGCAAGGTGAGCGAAGAACTAACCCATCTGGAGAACTGTCAAGGCGCGAGTGTG	420
Db	361	GGCGACAAAGAGGTGAGCGAAGAACTAACCCATCTGGAGAACTGTGAAGGCGCGAGTGTG	420
Qy	421	CACCAAGGCATCAGCCCCCGCACCTTGAACCGCTTGGTGAAGTATCGAGAGAGAGCC	480
Db	421	CACCAAGGCATCAGCCCCCGCACCTTGAACCGCTTGGTGAAGTATCGAGAGAGAGGCC	480
Qy	481	TTCAAGCCCGGAGGTATCCCACTGTTCAACCGGCTTGAGCGAGAGGCGCACCCCCAGAAC	540
Db	481	TTCAAGCCCGGAGGTATCCCACTGTTCAACCGGCTTGAGCGAGAGGCGCACCCCCAGAAC	540
Qy	541	CTGGAACAGATGTTGAACACCGTGGGCGGCAACAAGGCGGCATCAATGCTGAAGAC	600
Db	541	CTGGAACAGATGTTGAACACCGTGGGCGGCAACAAGGCGGCATCAATGCTGAAGAAC	600
Qy	601	ACCATCAACGAGAGAGCGCGGAGTGGGACCGGCTTGACCCCTGTGAGAGCGGCGCCGTG	660
Db	601	ACCATCAACGAGAGAGCGCGGAGTGGGACCGGCTTGACCCCTGTGACCGCGGCGCCGTG	660
Qy	661	GGCCCCGGGACAGATCGCGACCCCCCGGCGAGCGACATCGCGCGGCAACAAGACACTTG	720
Db	661	GGCCCCGGGACAGATCGCGAGCCCCCGGCGAGCGACATCGCGCGGCAACAAGACACTTG	720
Qy	721	CAGAGACAGATGCGCTTGATGACAGAACCCCCCGTGGCTGTGGAGCAATCTTAAG	780
Db	721	CAGAGACAGATGCGCTTGATGACAGAACCCCCCGTGTGGAGCAATCTTAAG	780
Qy	781	CGGTGATCATCTTGGGCTGTAACAAGATCGTGCGGATGTACGCCCCGTGAGCATCTGT	840
Db	781	CGGTGATCATCTTGGGCTGTAACAAGATCGTGCGGATGTACAGCCCCGTGAGCATCTGT	840
Qy	841	GACATCGGACAGGCGCCCAAGAGGCTTCCGCGCATACGTGACCGCTTCTTAAGACC	900
Db	841	GACATCAAGCAGGCGCCCAAGAGGCTTCCGCGCATACGTGACCGCTTCTTAAGACC	900
Qy	901	CTGGCGCGCGAGCGAGCCACCAAGAGGTGAAGAACTGGAAGACCGACACTCTCTGGTG	960
Db	901	CTGGCGCGCGAGCGAGCACTCAAGAGGTGAAGAACTGGAAGACCGACACTCTCTGGTG	960
Qy	961	CAGAACGCCAACCCCGACTGCAAAACCATCTTGCGCGCTTGCGCCCGCGGCGCAACCTG	1020
Db	961	CAGAACGCCAACCCCGACTGCAAAACCATCTTGCGCGCGCTTGCGCCCGCGGCGCAACCTG	1020
Qy	1021	GAGGAGATGATGACCGGCTTGCAAGGCGTGGGCGGCGCCCGCAAAAGCCCGTGTCTG	1080
Db	1021	GAGGAGATGATGACCGGCTTGCAAGGCGTGGGCGGCGCCCGCAAAAGCCCGTGTCTG	1080
Qy	1081	GCCGAGGCGATGAGCGAGGCGCAACAGCGTGAACATCATGTGCAAGAGCACTTCAAG	1140

Db 1081 GCCGAGCCATGAGCCAGCCCAACACCGAGCTGATGATCCAGAAAGCACTTCAAG 1140  
QY 1141 GGGCCCCGGCGCAAGTTCAAGTCTTCACTGGGCGAAGAGGCCCACTCCGCAAGAAC 1200  
Db 1141 GGGCCCCGGCGCGCTGTGAAGTCTTCACTGGGCGCGGAGGGCCCACTCCGCGCAAG 1200  
QY 1201 TGGCGGCGCGCGCGCAAGAGGGCTGTGAAGTGGGCGAAGAGGGCGCCACGATGAAG 1260  
Db 1201 TGGCGGCGCGCGCGCAAGCGCGCTGTGAAGTGGGCGAAGAGGGCCACGATGAAG 1260  
QY 1261 GACTGACCGAGCGCGCGCAACTTCTGGGCGAAGATTTGGCCACGCAAGAGGCGCG 1320  
Db 1261 GACTGACCGAGCGCGCGCAACTTCTGGGCGAAGATTTGGCCACGCAAGAGGCGCG 1320  
QY 1321 CCGCGCAATTTCTGTGAAGAACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377  
Db 1321 CCGCGCAATTTCTGTGAAGAACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
QY 1378 CCGCGCGCGCGAGCTTCCGCTTGTGAAGAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1437  
Db 1381 CCGCGCGCGCGAGCTTCAAGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431  
QY 1438 GACCGCGAGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1497  
Db 1432 G-----ACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1479  
QY 1498 CTGAGCGCACTAA 1509  
Db 1480 CTGAGCGCACTAA 1491

RESULT 13  
AAAS1609  
ID AAAS1609 standard; DNA; 1479 BP.  
AC AAAS1609;  
XX  
DT 31-OCT-2000 (first entry)  
XX  
DE HIV synthetic Gag polynucleotide.  
XX  
KM Gag; expression cassette; antigenic; type C; HIV; Env; synthetic;  
XX DNA immunization; packaging cell line; antigen presentation; ss.  
OS Human immunodeficiency virus; type C strain AF110965.  
OS Synthetic.  
FH Key Location/Qualifiers  
FT CDS 1..1479  
FT /tag= "a  
FT /product= "Synthetic\_Gag"  
FT /note= "Codon usage pattern was modified and inhibitory  
FT elements (INS) and RRE sites were inactivated resulting  
FT in improved expression"  
XX  
PN WO200039304-A2.  
PD 06-JUL-2000.  
XX  
PF 30-DEC-1999; 99WO-US031273.  
XX  
PR 31-DEC-1998; 98US-0114495P.  
PR 01-SEP-1999; 99US-0152195P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX Barnett S, Zur Megede J;  
PI  
DR MPI; 2000-452401/39.  
DR P-PSDB; AAY96943.  
XX  
PT Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env  
PT polypeptide and the polypeptide useful for immunizing a mammal especially

PT human against HIV.  
PS  
XX Claim 2, Page 92-93; 113pp; English.  
XX  
CC Expression cassettes comprising a polynucleotide encoding antigenic type  
CC C human immunodeficiency virus (HIV) Gag or Env polypeptides are useful  
CC in DNA immunization, generation of packaging cell lines and production of  
CC Gag- and/or Env-containing proteins. Synthetic Env and Gag expression  
CC cassettes exhibit increased potency for induction of cytotoxic T-  
CC lymphocyte (CTL) responses by DNA immunization. Gag of HIV-1 self-  
CC assemble into non-infectious virus-like particles which are used as a  
CC matrix for the proper presentation of an antigen entrapped or associated  
CC to the immune system of the host  
XX  
SQ Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;  
Query Match 84.6%; Score 1276.8; DB 3; Length 1479;  
Best Local Similarity 92.5%; Pred. No. 1.7e-148;  
Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;  
QY 1 ATGGGCGCGCGCGCGCGCACTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
QY 61 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 61 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
QY 121 CTGGAGGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
Db 121 CTGGAGGAAGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
QY 181 ATGAAGCAGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 181 ATCCGCGAGCTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
QY 241 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 241 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
QY 301 CTGGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 301 CTGGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCGGAC---GGCAGAGTGAAGCGCAACTACCTGCTGCGAGAGAGAGAGAGAGAGAG 417  
Db 361 GCGGAC---GGCAGAGTGAAGCGCAACTACCTGCTGCGAGAGAGAGAGAGAGAGAG 420  
QY 418 GTGCAACAGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 477  
Db 421 GTGCAACAGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
QY 478 GCGTTACGCGCGCGAGTATCCCATGTTCAACCGCGCTGAGAGAGAGAGAGAGAGAG 537  
Db 481 GCGTTACGCGCGCGAGTATCCCATGTTCAACCGCGCTGAGAGAGAGAGAGAGAGAG 540  
QY 538 GACTGAACAGATGTTGAACAACGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597  
Db 541 GACTGAACAGATGTTGAACAACGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
QY 598 GACACCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657  
Db 601 GACACCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 658 GTGGCGCGCGCGCGAGTGGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717  
Db 661 ATGCGCGCGCGCGAGATGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
QY 718 CTGCAAGAGCAGATGCTGTGATGACGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 777  
Db 721 CTGCAAGAGCAGATGCTGTGATGACGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
QY 778 AAGCGGTGATCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 837

Db	781	AAGCGTGGATCTATCTTGGGCTCTGAACAGATGTGCGAGTGAACGCCCCGAGGATC	840
Qy	838	CTGGACATCCGCGCAGGCCCCCAAGAGACCCTTCGCGACTACGTGACCGCTCTTCAAG	897
Db	841	CTGGACATCAAGCAGGAGCCCCCAAGAGACCTTCCGCGACTACGTGACCGCTCTTCAAG	900
Qy	898	ACCTTCGGGCGGAGGCGCAACCCAGAGACGTAAAGACTGATGACCGAGACCTTCGCTG	957
Db	901	ACCTTCGGGCGGAGGCGCAACCCAGAGAGTAAAGACTGATGACCGAGACCTTCGCTG	960
Qy	958	GTGCAGAACGCCAACCCCGCACTGCAAGACCATCTCGCGCTTCGAGCCCCGCGCACCC	101
Db	961	GTGCAGAAACGCCAACCCCGCACTGCAAGACCATCTCGCGCTTCGAGCCCCGCGCACCC	102
Qy	1018	CTGGAGGAATATGATACCGCTTCGCAAGGCGTGTGGCGGCCCGCGCCACAAGCCCGCTG	107
Db	1021	CTGGAGGAATATGATACCGCTTCGCAAGGCGGTGTGGCGGCCCGCGCCACAAGCCCGCTG	108
Qy	1078	CTGGCGGAGGCGATGAGCGAGGCCAACAGGTGAAATCATATATGCAAGAGCACTTC	113
Db	1081	CTGGCGGAGGCGATGAGCGAGGCCAACAGGTGAAATCATATATGCAAGAGCACTTC	113
Qy	1138	AAGGCGCCCCCGCGCAACGTCAAGTGTCTTCAACTGCGCGCAAGAGGCGCACTGCGCAAG	119
Db	1138	AAGGCGCCCCCGCGCGCATGTCAAGTGTCTTCAACTGCGCGCAAGAGGCGCACTGCGCGCG	119
Qy	1138	AACCTCCCGGCCCCCGCGCAAGAGGCGCTGTGAAGTGTGGCGCAAGAGGCGCAACAGATG	125
Db	1138	AACCTCCCGGCCCCCGCGCAAGAGGCGCTGTGAAGTGTGGCGCAAGAGGCGCAACAGATG	125
Qy	1258	AAGACTGTGACCGAGCGCGCAGGCGCAACTTCTGTGGCGAATCTGGCGCCAGCCACAAGGCG	131
Db	1258	AAGACTGTGACCGAGCGCGCAGGCGCAACTTCTGTGGCGAATCTGGCGCCAGCCACAAGGCG	131
Qy	1318	CGCCCGCGCACTTCTCTGAGAACCGCGACGAGCCCGCGCCCCCAACCGTGTCCACCGCG	137
Db	1318	CGCCCGCGCACTTCTCTGAGAGCGCG-----CCCGAGGCCACCGCG	135
Qy	1378	CCCCCGCGGAGAGGTCTTCGCTTCGAGAGAACACACCCCGCGCCCAAGCAGAGCGCCAAAG	143
Db	1360	CCCCCGCGGAGAGGTCTTCGCTTCGAGAGAACACACCCCGCGCCCAAGCAGAGCGCCAAAG	141
Qy	1438	GACCGCGAGCCTTACCGCGAGCGCCCTGTACCGCGCCGTGGCGAGCTGTTGGGACGGGCGCC	149
Db	1420	G-----ACCGCGAGACCTGTACCGCGCTGAAGAGCCTGTTGGGACAGACCC	146
Qy	1498	CTGAGCCACTTAA	1509
Db	1468	CTGAGCCACTTAA	1479

RESULT 14	
XX AAL44548	
ID AAL44548	standard, DNA, 1479 BP.
XX AC	
XX AAL44548;	
XX DT	29-AUG-2003 (revised)
XX DT	08-NOV-2002 (first entry)
XX DE	HIV-1 p55gag polypeptide coding sequence 1.
XX KW	HIV; ds; vaccine; gene; immune response; microparticle;
XX KW	absorbent surfact; poly(alpha-hydroxy acid); polyhydroxy butyric acid;
XX KW	submicron emulsion; viral infection; bacterial infection;
XX KW	parasitic infection; HIV-1 p55gag polypeptide.
OS	Human immunodeficiency virus 1.
XX	
PN	WO200226209-A2.

Query Match	84.6%	Score 1276.8	DB 6	Length 1479
Best Local Similarity	92.5%	Pred. No. 1.7e-148		
Matches 1399	Conservative 0	Mismatches 77	Indels 36	Gaps 4
QY	1	ATGGGCGCGCCCGGCGCAGCATCTGGCGCGCGGCGGAGAGCTGGACAAAGTGGAGAAATCCGC	60	
DB	1	ATGGGCGCGCCCGGCGCAGCATCTGGCGCGCGGCGGAGAGCTGGACAAAGTGGACATCCGC	60	
QY	61	CTGCGCGCCCGGCGGAGAAAGCACTACATGCTGAAGCACTGGTGTGGGCGACCGCGAG	120	
DB	61	CTGCGCGCCCGGCGGAGAAAGTCTACATGATGAAGCACTGGTGTGGGCGACCGCGAG	120	
QY	121	CTGAGAGGCGTTCGCTCTGAATCCCGCGCTCTGTAGAGACCGCGGAGGCTGCAGAGATC	180	
DB	121	CTGAGAGAGTTTCGCTCTGAATCCCGCGCTCTGTAGAGACCGCGGAGGCTGCAGAGATC	180	
QY	181	ATGAAGCAGCTGACGCGCGCCCTGTAGACCGGACCGGAGAGCTGGCGAGCTGTACAA	240	
DB	181	ATTCGCGCAGCTGACCGCGCCCTGTAGACCGGACCGGAGAGCTGTAAAGCTGTTCAC	240	
QY	241	ACCGTGGCACCCTGTACTGTGCGTGAACCGCGCATCGAGTCCGCGACACCAAGAGGCC	300	
DB	241	ACCGTGGCACCCTGTACTGTGCGTGAACGAGAAATCGAGTCCGCGACACCAAGAGGCC	300	
QY	301	CTGAGCAAGATCGAGAGAGAGAGCAAGATGCCAGAGAAAGCCCGACAGGCCAAGAG	360	
DB	301	CTGAGCAAGATCGAGAGAGAGAGCAAGATGCCAGAGAAATTCAGAGGCCAGAGGCC	360	
QY	361	GCCGAC---GGCAAGTGAAGCCAGAACTAACCCATCGGAGAACTGCAAGGCGCAGATG	417	
DB	361	GCCGACAAAGGCGAGTGAAGCCAGAACTAACCCATCGTGAAGAACTTGCAAGGCGCAGATG	420	
QY	418	GTGACCAAGGCGCATCAAGCCCGCAGACCTGTAAACCTGTGGGTGAAGGTGATCGAGAGAG	477	
DB	421	GTGACCAAGGCGCATCAAGCCCGCAGACCTGTAAACCTGTGGGTGAAGGTGATGAGAGAGAG	480	
QY	478	GCCTTAAAGCCCGAGGTGATTCCTCATGTTCAACCGCCTTAAGCGAGGCGCAGCCCGCAG	537	

481 GCCTTCAGCCCGAGGATGATCCCATGTTCAACCGCCCTGAGCGAGGCGCCACCCCCAG 540  
QY 538 GACCTGAACCAATGTTGAACACCTGGGGGCGCCACCAAGGCCCTTCAGATGCTGAAG 597  
QY 541 GACCTGAACCAATGTTGAACACCTGGGGGCGCCACCAAGGCCCTTCAGATGCTGAAG 600  
QY 598 GACACCATCAAGAGAGGCGCGGAGTGGAGCCGCTGACCCGCTGACGCGCGGCGCC 657  
DB 601 GACACCATCAAGAGAGGCGCGGAGTGGAGCCGCTGACCCGCTGACGCGCGGCGCC 660  
QY 658 GTGGCCCCCGGCGAGATGGCGGACCCCGCGGCGAGCACTCGCGCGCGCCACAGCACC 717  
DB 661 ATCGCCCCCGGCGAGATGGCGGAGCCCGCGGCGAGCACTCGCGCGCGCCACAGCACC 720  
QY 718 CTGCGAGAGCAATGCTGCTGATGACCAAGCAACCCCGCTGCGCGCGGCGCATCTAC 777  
DB 721 CTGCGAGAGCAATGCTGCTGATGACCAAGCAACCCCGCTGCGCGCGGCGCATCTAC 780  
QY 778 AAGCGGTGATCATCTGGGCGCTGAACAAAGATGTGCGGATGTACAGCCCGCTGAGCATC 837  
DB 781 AAGCGGTGATCATCTGGGCGCTGAACAAAGATGTGCGGATGTACAGCCCGCTGAGCATC 840  
QY 838 CTGCAATCCGCGAGGCGCCCAAGAGAGCCCTTCGCGACTACCTGAGCGCTTCTTCAAG 897  
DB 841 CTGCAATCAAGAGGCGCCCAAGAGAGCCCTTCGCGACTACCTGAGCGCTTCTTCAAG 900  
QY 898 ACCCTGCGCGCGAGAGGCGCCACCAAGAGATGTGAAGACTGTGATGACCAAGCCTGCTG 957  
DB 901 ACCCTGCGCGCGAGAGGCGCCACCAAGAGATGTGAAGACTGTGATGACCAAGCCTGCTG 960  
QY 958 GTGCGAAGAGCCCAACCCGAGCTGCAAGACATCTGCGGCGCTGCGGCGCGCGCGCAGC 1017  
DB 961 GTGCGAAGAGCCCAACCCGAGCTGCAAGACATCTGCGGCGCTGCGGCGCGCGCGCAGC 1020  
QY 1018 CTGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCGCGCCACAGGCGCGCTG 1077  
DB 1021 CTGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCGCGCCACAGGCGCGCTG 1080  
QY 1078 CTGGCGGAGCGATGAGCGCCCAAGCGCTGATGATGATGCAAGAGCAATTC 1137  
DB 1081 CTGGCGGAGCGATGAGCGCCCAAGCGCTGATGATGATGCAAGAGCAATTC 1137  
QY 1138 AAGGGCGCGCGGCGCAAGTGTCTCACTGCGGCGAGAGGCGCCACATCGCCAG 1197  
DB 1138 AAGGGCGCGCGGCGCAAGTGTCTCACTGCGGCGAGAGGCGCCACATCGCCAG 1197  
QY 1198 AACTGCGCGCGCGCGCAAGAGGCTGCTGGAATGCGCGCAAGAGGCGCCACAGATG 1257  
DB 1198 AACTGCGCGCGCGCGCGCAAGAGGCTGCTGGAATGCGCGCAAGAGGCGCCACAGATG 1257  
QY 1258 AAGAGCTGACCGAGCGCGCAAGCTTCTGCGCAAGATCTGCGCCACAGAGGCG 1317  
DB 1258 AAGAGCTGACCGAGCGCGCGCAAGCTTCTGCGCAAGATCTGCGCCACAGAGGCG 1317  
QY 1318 CGCGCGCGCACTCTGCAAGAACCGCAAGAGCCCGCGCGCCCGCGCGCGCCACCGCG 1377  
DB 1318 CGCGCGCGCACTCTGCAAGAGCGC-----CGCGCGCGCCACCGCGCG 1359  
QY 1378 CGCGCGCGCGAGCTTCCGCTTCAAGAGACCAACCGCGCGCCCAAGAGAGGCGCGAG 1437  
DB 1360 CGCGCGCGCGAGAGCTTCCGCTTCAAGAGACCAACCGCGCGCGCAAGAGAGAGAG 1419  
QY 1438 GACCGGAGCGCTTCCGCGAGCGCGCTGACCGCGCTGCGAGCGCTGTTGCGAGCGCGCC 1497  
DB 1420 G-----ACCGGAGAGCGCTGACCGAGCTGAGAGAGCGCTGTTGCGGCAAGCGAGCC 1467  
QY 1498 CTGAGCGAGTAA 1509  
DB 1468 CTGAGCGAGTAA 1479

ID ABL39953 standard; DNA; 1479 BP.  
XX  
AC ABL39953;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Synthetic Gag polynucleotide sequence SEQ ID NO:3.  
XX  
KW Human immunodeficiency virus type C; antigenic HIV type C protein;  
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
KW immunostimulant; gene therapy; gene; ds.  
OS Human immunodeficiency virus; type C.  
OS Synthetic.  
PN WO200204493-A2.  
PN 17-JAN-2002.  
PD  
PF 05-JUL-2001; 2001WO-US021241.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
PA (CHIR) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megele J, Barnett SM, Engelbrecht S, Van Rensburg EJ;  
XX WPI, 2002-154920/20.  
DR  
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
PT applications including DNA immunization or generation of packaging cell  
PT lines, particularly in gene therapy.  
XX  
PS Example 1; Fig 1; 233p; English.  
XX  
CC The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef  
CC (I). (i) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;  
Query Match 84.6%; Score 1276.8; DB 6; Length 1479;  
Best Local Similarity 92.5%; Pred. No. 1.7e-148;  
Matches 1399; Conservative 0; Mismatches 77; Indels 36; Gaps 4;  
QY 1 AATGGGCGCGCGCGCGAGATCTGCGCGCGCGAGAGCTGGAACAATGGGAGAAATCCGC 60  
DB 1 AATGGGCGCGCGCGCGAGATCTGCGCGCGCGAGAGCTGGAAGCTTGGGAGGCAATCCGC 60  
QY 61 CTGGCGCGCGCGCGAGAGCAATGCTGGAAGCACTGCTGCGCGCGAGCGCGCGAG 120  
DB 61 CTGGCGCGCGCGCGAGAGAGTGTATGATGAGCACTGCTGCGCGCGAGCGCGAG 120  
QY 121 CTGAGAGGCTTCCGCTGAAACCCGCGCTGCTGAGACCGCGAGGCGCTGCAAGCATC 180  
DB 121 CTGAGAGGCTTCCGCTGAAACCCGCGCTGCTGAGACCGCGAGGCGCTGCAAGCATC 180  
QY 181 ATGAAGAGCTGCAAGCGCGCGCTGCAAGACCGGCAACCGAGAGCTGCAAGCATC 240  
DB 181 ATTCGCGAGCTGCAAGCGCGCGCTGCAAGACCGGCAACCGAGAGCTGCAAGCATC 240  
QY 241 ACCGTGCGCACCTGCTGATGCTGCGTGCACCGCGCATGAGTCCGCGAGCAACAGAGGCGC 300  
DB 241 ACCGTGCGCACCTGCTGATGCTGCGTGCACCGAGAGATGAGTCCGCGAGCAACAGAGGCGC 300

QY 301 CTGGAAGATCGAGGAGGAGCAAGAACTCCCAAGTAAAGACCCAGCGCCCAAGAG 360  
 Db 301 CTGGAAGATCGAGGAGGAGCAAGAACTCCCAAGTAAAGACCCAGCGCCCAAGAG 360  
 QY 361 GCCGAC---GGCAAGGTGAGCCAGAACTACCCCATCTGAGAGACCTGCAAGGCGCAATG 417  
 Db 361 GCCGACAGAGGAGCAAGGTGAGCCAGAACTACCCCATCTGAGAGACCTGCAAGGCGCAATG 420  
 QY 418 GTGACACCAAGGCGCATCAGCCCCCGACCTTGAACGCTGTGGGTGAAGGTGATCGAGAGAG 477  
 Db 421 GTGACACCAAGGCGCATCAGCCCCCGACCTTGAACGCTGTGGGTGAAGGTGATCGAGAGAG 480  
 QY 478 GCGTTAGCCCCGAGGAGTATCCCATGTTTCAACGCGCTGAGCGAGAGGCGCCAGCCCCAG 537  
 Db 481 GCGTTAGCCCCGAGGAGTATCCCATGTTTCAACGCGCTGAGCGAGAGGCGCCAGCCCCAG 540  
 QY 538 GACCTGAACAAGATGTTGAACAACGTTGGGCGGCGCACAGGCGCGCATGCGAGATGCTGAAG 597  
 Db 541 GACCTGAACAAGATGTTGAACAACGTTGGGCGGCGCACAGGCGCGCATGCGAGATGCTGAAG 600  
 QY 598 GACACCATCAACAGAGAGGCGCGCGAGTGGGACCGCTGCAACCCGTGCAAGCGCGGCC 657  
 Db 601 GACACCATCAACAGAGAGGCGCGCGAGTGGGACCGCTGCAACCCGTGCAAGCGCGGCC 660  
 QY 658 GTGCGCCCCGAGGAGTGGCGCGACCCCGCGGAGGAGCATGCGCGCGCGCACAGCAACC 717  
 Db 661 ATGCGCCCCGAGGAGTGGCGCGACCCCGCGGAGGAGCATGCGCGCGCGCACAGCAACC 720  
 QY 718 CTGCAAGAGCAAGATGCGCTGATGACCAAGACCCCGGTGCGCTGGGCGACATCTTAC 777  
 Db 721 CTGCAAGAGCAAGATGCGCTGATGACCAAGACCCCGGTGCGCTGGGCGACATCTTAC 780  
 QY 778 AAGCGGTGATCATCTTGGGCGCTGAAACAAGATGTGCGGATGTACAGCCCCGTGAGCATC 837  
 Db 781 AAGCGGTGATCATCTTGGGCGCTGAAACAAGATGTGCGGATGTACAGCCCCGTGAGCATC 840  
 QY 838 CTGGAATATCGGCGAGGCGCCCAAGAGGCGCTTCCGCGACCTACGTTGAGCGCTTCAAG 897  
 Db 841 CTGGAATATCGGCGAGGCGCCCAAGAGGCGCTTCCGCGACCTACGTTGAGCGCTTCAAG 900  
 QY 898 ACCCTGCGCGCGAGGAGCGCAACCCAGAGCTGTAAGAACTGATGACCGAGACCTGTCTG 957  
 Db 901 ACCCTGCGCGCGAGGAGCGCAACCCAGAGCTGTAAGAACTGATGACCGAGACCTGTCTG 960  
 QY 958 GTGCAAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCCCGCGCCAGC 1017  
 Db 961 GTGCAAGAACGCCAACCCCGACTGCAAGACCATCTGCGCGCTCTGCGCCCCCGCGCCAGC 1020  
 QY 1018 CTGGAAGAGATGATGACCGCTGCGAGGCGCTGCGGCGCGCCGCGCACAAAGCGCGCTG 1077  
 Db 1021 CTGGAAGAGATGATGACCGCTGCGAGGCGCTGCGGCGCGCCGCGCACAAAGCGCGCTG 1080  
 QY 1078 CTGCGCGAGGCGATGAGCGCGCAACAAGCTGAACATCATGATGCAAGAGCAACTTC 1137  
 Db 1081 CTGCGCGAGGCGATGAGCGCGCAACAAC---AGCGTATGATGCAAGAGCAACTTC 1137  
 QY 1138 AAGGCGCCCCGCGGCAAGCTCAAGTCTTCAACTGCGCAAGAGGCGCACATGCGCCAG 1197  
 Db 1138 AAGGCGCCCCGCGGCAAGTCTTCAAGTCTTCAACTGCGCAAGAGGCGCACATGCGCCAG 1197  
 QY 1198 AACTGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGGGCAAGAGGCGCACAGATG 1257  
 Db 1258 AAGGCTGCAACCGAGCGCGCAACTTCTGTGGCAAGATGTGGCCCAAGGCGC 1317  
 QY 1258 AAGGCTGCAACCGAGCGCGCAACTTCTGTGGCAAGATGTGGCCCAAGGCGC 1317  
 Db 1258 AAGGCTGCAACCGAGCGCGCAACTTCTGTGGCAAGATGTGGCCCAAGGCGC 1317  
 QY 1318 CGCGCGCGCAACTTCTGTGAGAACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377  
 Db 1318 CGCGCGCGCAACTTCTGTGAGAACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1359

QY 1378 CCCCCCGCGAGAGCTTCCGCTTTCAGAGAGACCAACCCCGCGCCCAAGCGAGAGCCCAAG 1437  
 Db 1360 CCCCCCGCGAGAGCTTCCGCTTTCAGAGAGAGACCAACCCCGCGCCCAAGCGAGAGCAAG 1419  
 QY 1438 GACCGGAGGCGCTTACCGGAGGCGCGCTGACCGCGCTGCGGAGCGCTGTTGGCGAGCGGCGCC 1497  
 Db 1420 G-----ACCGGAGGCGCTTACCGGAGGCGCGCTGAGCGCTGAGAGGCGCTTGGGAGAGCGCC 1467  
 QY 1498 CTGAGCGAGTAA 1509  
 Db 1468 CTGAGCGAGTAA 1479

Search completed: March 12, 2005, 17:35:57  
 Job time : 862.019 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 13:16:01 ; Search time 6734.44 Seconds  
(without alignments)  
10857.457 Million cell updates/second

Title: US-09-475-704A-4

Perfect score: 1509  
Sequence: 1 atggcgccgcgcacat.....gcgccccctgagccagtaa 1509

Scoring table: IDENTITY\_NUC

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

### Listing first 45 summaries

Database : GenEmbl : \*

```

1:  gb_ba:*
2:  gb_hcg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_stb:*
12: gb_un:*
13: gb_yn:*
14: gb_vi:*

```

**Pred.** No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1509	100.0	1509	6	AX455888	AX455888 Sequence
2	1507.4	99.9	1509	6	AX468544	AX468544 Sequence
3	1504.4	98.7	1509	6	AX468548	AX468548 Sequence
4	1489.8	99.7	1509	6	AX455905	AX455905 Sequence
5	1297.6	86.0	1494	6	AX455935	AX455935 Sequence
6	1294.4	85.8	1491	6	AX455983	AX455983 Sequence
7	1291.4	85.6	1485	12	AY181195	AY181195 Synthyrid
8	1276.8	84.6	1479	6	AX455887	AX455887 Sequence
9	1276.8	84.6	1479	6	AX468543	AX468543 Sequence
10	1271.2	84.2	1479	6	AX468547	AX468547 Sequence
11	1260.8	83.6	1479	6	AX455904	AX455904 Sequence
12	1202.2	79.7	4288	6	AX149648	AX149648 Sequence
13	1201	79.6	1509	12	AF201927	AF201927 Synthyrid
14	1201	79.6	1515	6	BD63637	BD63637 Improved
15	1201	79.6	1515	6	CO870496	CO870496 Sequence
16	1201	79.6	1515	6	AR371320	AR371320 Sequence
17	1201	79.6	4472	6	BD63697	BD63697 Improved
18	1201	79.6	4472	6	CO870567	CO870567 Sequence
19	1201	79.6	4472	6	AR373380	AR373380 Sequence

20	1201	79.6	4608	6	BD263658	BD263658	Improved
21	1201	79.6	4608	6	C0870568	C0870568	Sequence
22	1201	79.6	4608	6	AR373381	AR373381	Sequence
23	1201	79.6	4689	6	BD263656	BD263656	Improved
24	1201	79.6	4689	6	C0870566	C0870566	Improved
25	1201	79.6	4689	6	AR373379	AR373379	Sequence
26	1201	79.6	4766	6	BD263655	BD263655	Improved
27	1201	79.6	4766	6	C0870565	C0870565	Sequence
28	1201	79.6	4766	6	AR373378	AR373378	Sequence
29	1199	79.5	2031	6	BD263640	BD263640	Improved
30	1199	79.5	2031	6	C0870499	C0870499	Sequence
31	1199	79.5	2031	6	AR373323	AR373323	Sequence
32	1195.8	79.2	1503	12	AY53163	AY53163	Synthetic
33	1186	78.6	9166	6	AX427930	AX427930	Sequence
34	1170.8	77.6	9788	6	AX427936	AX427936	Sequence
35	1145	75.9	1539	6	AX306431	AX306431	Sequence
36	1143.2	75.8	1548	12	AF287354	AF287354	Synthetic
37	1119	74.2	6438	6	AX427919	AX427919	Sequence
38	1119	74.2	9194	6	AX427925	AX427925	Sequence
39	1119	74.2	9194	6	AX427926	AX427926	Sequence
40	1119	74.2	12411	6	AX427927	AX427927	Sequence
41	1117.8	74.1	1847	12	AF202464	AF202464	Synthetic
42	1117.8	74.1	1847	12	AF202465	AF202465	Synthetic
43	1117.8	74.1	1853	6	BD263638	BD263638	Improved
44	1117.8	74.1	1853	6	C0870497	C0870497	Sequence
45	1117.8	74.1	1853	6	AR373321	AR373321	Sequence

## ALIGNMENTS

RESULT 1	AX455888	1509 bp	DNA	linear	PAT 06-JUL-2002
LOCUS	AX455888	1509 bp	DNA	linear	PAT 06-JUL-2002
DEFINITION	Sequence 4 from Patent WO0204493.				
ACCESSION	AX455888				
VERSION	AX455888.1	GI:21714881			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE	synthetic construct				
AUTHORS	synthetic construct				
TITLE	other sequences; artificial sequences.				
JOURNAL	zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E. Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof Patent: WO 0204493-A 4 17-JAN-2002; CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)				
FEATURES	Location/Qualifiers				
source	1..1509				
	/organism="synthetic construct"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:32630"				
ORIGIN	/note="synthetic Gag of HIV strain AF110967"				
Query Match	100.0%;	Score 1509;	DB 6;	Length 1509;	
Best Local Similarity	100.0%;	Pred. No. 3.2e-150;			
Matches 1509;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	ATGGGCGCCCGCGCAGCATCTGCGCGCGGAGAGCTGGACAAGTGGAGAAATCCGC	60		
Db	1	ATGGGCGCCCGCGCGCAGCATCTGCGCGGAGAGAGCTGGACAAGTGGAGAAATCCGC	60		
QY	61	CTGGCGCCCGCGCGCAAGACACTACATGCTGAAGCACTGTGTGGCGCAGCCCGAG	120		
Db	61	CTGGCGCCCGCGCGCAAGACACTACATGCTGAAGCACTGTGTGGCGCAGCCCGAG	120		
QY	121	CTGAGAGGGCTTTCGCGCTGGAACCCCGGCTCTGTGAAGACCGCCGAGGGCTCTGAAGCATC	180		
Db	121	CTGAGAGGGCTTTCGCGCTGGAACCCCGGCTCTGTGAAGACCGCCGAGGGCTCTGAAGCATC	180		
QY	181	ATGAAGAGCTTGACGCCGCGCTTGACAGCCGACCGAGGAGCTGCCAGCTGTACAC	240		



Db 181 ATGAAGACGTGACGCCGCCCTGACAGACCGGAGAGCTGGACGCTGTACAC 240  
 QY 241 ACCGTGGCCACCTGTACTGCTGTGACGCCGGATGAGGTCCGCAACCAAGAGCC 300  
 Db 241 ACCGTGGCCACCTGTACTGCTGTGACGCCGGATGAGGTCCGCAACCAAGAGCC 300  
 QY 301 CTGGAACAAGATCGAAGAGAGAGCAAGATCCACAGAGACCCAGAGGCGAAGAG 360  
 Db 301 CTGGAACAAGATCGAAGAGAGAGCAAGATCCACAGAGACCCAGAGGCGAAGAG 360  
 QY 361 GCCGACGCAAGATGAGCCAGAACTACCCCATGTGCAAGACTTGAGGCGCAGATG 420  
 Db 361 GCCGACGCAAGATGAGCCAGAACTACCCCATGTGCAAGACTTGAGGCGCAGATG 420  
 QY 421 CACCGAGGCATGACGCCCGGACCCCTGAAAGCTGTGGTGAAGGTATCGAAGAGG 480  
 Db 421 CACCGAGGCATGACGCCCGGACCCCTGAAAGCTGTGGTGAAGGTATCGAAGAGG 480  
 QY 481 TTCAAGCCCGAGGTGATCCCAATGTTCAACGCTGTGAGGAGGCGCACCCCGAG 540  
 Db 481 TTCAAGCCCGAGGTGATCCCAATGTTCAACGCTGTGAGGAGGCGCACCCCGAG 540  
 QY 541 CTGAACACGATGTTGAACACCGTGGGCGCCACAGGCGCCATGCAATGCTGAAG 600  
 Db 541 CTGAACACGATGTTGAACACCGTGGGCGCCACAGGCGCCATGCAATGCTGAAG 600  
 QY 601 ACCATCAAGAGAGGCGCGGAGTGGGACCGCTGTGCAACCCCTGTGAGGCGCG 660  
 Db 601 ACCATCAAGAGAGGCGCGGAGTGGGACCGCTGTGCAACCCCTGTGAGGCGCG 660  
 QY 661 GCCCGCGGACAGATGCGACGCCCGGCGAGCAGATGCGGCGGCGCCACAGACCC 720  
 Db 661 GCCCGCGGACAGATGCGACGCCCGGCGAGCAGATGCGGCGGCGCCACAGACCC 720  
 QY 721 CAGAGACGATGCTCTGATGACCAAGCAACCCCGCTGCTGGGCGCACTTCAAG 780  
 Db 721 CAGAGACGATGCTCTGATGACCAAGCAACCCCGCTGCTGGGCGCACTTCAAG 780  
 QY 781 CGGTGATCATCTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAAGCT 840  
 Db 781 CGGTGATCATCTGGGCGCTGAAACAAGATCGTGGGATGTACAGCCCGTGAAGCT 840  
 QY 841 GACATCCGCGACAGGCGCCCAAGAGGCGCTTCCGCACTAGTGAACCGCTTCTCA 900  
 Db 841 GACATCCGCGACAGGCGCCCAAGAGGCGCTTCCGCACTAGTGAACCGCTTCTCA 900  
 QY 901 CTGCGCGCGGACAGAGCCCAAGAGCTGGAAGAACTGATGACCGAACCCTGCTG 960  
 Db 901 CTGCGCGCGGACAGAGCCCAAGAGCTGGAAGAACTGATGACCGAACCCTGCTG 960  
 QY 961 CAGAAAGCAACCCGCACTGCAAGACCAATCTCGGCGCTCTGGCCCGGCGCACCT 1020  
 Db 961 CAGAAAGCAACCCGCACTGCAAGACCAATCTCGGCGCTCTGGCCCGGCGCACCT 1020  
 QY 1021 GAGGATGATGACCGCTGCGAGGCGTGGGCGGCGCCGCGCACAGGCGCGCTG 1080  
 Db 1021 GAGGATGATGACCGCTGCGAGGCGTGGGCGGCGCCGCGCACAGGCGCGCTG 1080  
 QY 1081 GCCGAGGAGATGAGCGCAGAGCCCAAGCGTGAACATCATGATGCAAGAGCACTT 1140  
 Db 1081 GCCGAGGAGATGAGCGCAGAGCCCAAGCGTGAACATCATGATGCAAGAGCACTT 1140  
 QY 1141 GGGCCCGCGGCAACGTCAGAGTCTTCAATGCGGCAAGAGAGGCAATGCGCCAG 1200  
 Db 1141 GGGCCCGCGGCAACGTCAGAGTCTTCAATGCGGCAAGAGAGGCAATGCGCCAG 1200  
 QY 1201 TGGCGCGCGCGCAAGAGGCGCTGGAAGTGGCAAGAGGCGCAACAGATGAG 1260  
 Db 1201 TGGCGCGCGCGCAAGAGGCGCTGGAAGTGGCAAGAGGCGCAACAGATGAG 1260  
 QY 1261 GACTGCAACGAGCGCAGAGCCCACTTCTGTGGCAAGATCTGCGCCAGCAAGGCG 1320  
 Db 1261 GACTGCAACGAGCGCAGAGCCCACTTCTGTGGCAAGATCTGCGCCAGCAAGGCG 1320

QY 1321 CCGGCAACTTCTGCAAGACCGGAGCGAGCCCGCGCCCGCCACCGTGCCACCGCC 1380  
 Db 1321 CCGGCAACTTCTGCAAGACCGGAGCGAGCCCGCGCCCGCCACCGTGCCACCGCC 1380  
 QY 1381 CCGGCGAGGCTTCCGCTTGAAGAGACCAACCCCGCGCCCAAGAGAGCCCAAG 1440  
 Db 1381 CCGGCGAGGCTTCCGCTTGAAGAGACCAACCCCGCGCCCAAGAGAGCCCAAG 1440  
 QY 1441 CCGGAGCCCTACCGGAGCCCGCTGACCGCCCTGCGCAGCCTGTTGCGAGCGCC 1500  
 Db 1441 CCGGAGCCCTACCGGAGCCCGCTGACCGCCCTGCGCAGCCTGTTGCGAGCGCC 1500  
 QY 1501 AGCCAGTAA 1509  
 Db 1501 AGCCAGTAA 1509  
 RESULT 2  
 AX468544  
 LOCUS AX468544 1509 bp DNA linear PAT 16-JUL-2002  
 DEFINITION Sequence 64 from Patent WO0226209.  
 ACCESSION AX468544  
 VERSION AX468544.1 GI:21901374  
 KEYWORDS  
 SOURCE Human immunodeficiency virus 1 (HIV-1)  
 ORGANISM Human immunodeficiency virus 1  
 Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate  
 lentivirus group.  
 REFERENCE  
 1 O'Hagan, D., Otten, G., Donnelly, J. J., Polo, J. M., Barnett, S.,  
 Singh, M., Uimer, J. and Dubensky, T. W.,  
 Microparticles for delivery of the heterologous nucleic acids  
 Patent: WO 0226209-A 64 04-APR-2002;  
 CHIRON CORPORATION (US)  
 FEATURES  
 source 1..1509  
 /organism="Human immunodeficiency virus 1"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:11676"  
 ORIGIN  
 Query Match 99.9%; Score 1507.4; DB 6; Length 1509;  
 Best Local Similarity 99.9%; Pred. No. 4.7e-150;  
 Matches 1508; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATGGGCGCCCGCGCAGACATCTGCGCGGAGAGAGCTGGAACAAGTGGAGATCCG 60  
 Db 1 ATGGGCGCCCGCGCAGACATCTGCGCGGAGAGAGCTGGAACAAGTGGAGATCCG 60  
 QY 61 CTGCGCCCGCGGCAAGAGCACTATGCTGAAGCACTGGTGTGGCCAGCCGCGAG 120  
 Db 61 CTGCGCCCGCGGCAAGAGCACTATGCTGAAGCACTGGTGTGGCCAGCCGCGAG 120  
 QY 121 CTGAGAGGCTTGGCCCTGAAACCCCGGCTGTGGAAGACCGCGAGGCTGCAAGCATC 180  
 Db 121 CTGAGAGGCTTGGCCCTGAAACCCCGGCTGTGGAAGACCGCGAGGCTGCAAGCATC 180  
 QY 181 ATGAAGAGCTGCAAGCCCGCTTGACAGCCGGAACCGAGAGGCTGCGACCTGTAC 240  
 Db 181 ATGAAGAGCTGCAAGCCCGCTTGACAGCCGGAACCGAGAGGCTGCGACCTGTAC 240  
 QY 241 ACCGTGGCCACCTGTACTGCTGTGACGCCGGATGAGGTCCGCAACCAAGAGGCC 300  
 Db 241 ACCGTGGCCACCTGTACTGCTGTGACGCCGGATGAGGTCCGCAACCAAGAGGCC 300  
 QY 301 CTGGAACAAGATCGAAGAGAGAGCAAGATCCACAGAGACCCAGAGGCGAAGAG 360  
 Db 301 CTGGAACAAGATCGAAGAGAGAGCAAGATCCACAGAGACCCAGAGGCGAAGAG 360  
 QY 361 GCCGACGCAAGATGAGCCAGAACTACCCCATGTGCAAGACTTGAGGCGCAGATG 420  
 Db 361 GCCGACGCAAGATGAGCCAGAACTACCCCATGTGCAAGACTTGAGGCGCAGATG 420



421 CACGAGGCGATGAGCCCGGCAACCTGAAAGCGCTGGGTGAAGTGTATCGAGAGAAAGGCC 480  
421 CACGAGGCGATGAGCCCGGCAACCTGAAAGCGCTGGGTGAAGTGTATCGAGAGAAAGGCC 480  
481 TTCAAGCCCGAGAGTGAATCCCATGTTTACCGCTCTGAGGAGAGGCGCCACCCCGCAGAGC 540  
481 TTCAAGCCCGAGAGTGAATCCCATGTTTACCGCTCTGAGGAGAGGCGCCACCCCGCAGAGC 540  
541 CTGAACACGATGTTGAACACCGTGGGCGGCGCACAGGCGCGCATGCGATGCTGAAGAGAC 600  
541 CTGAACACGATGTTGAACACCGTGGGCGGCGCACAGGCGCGCATGCGATGCTGAAGAGAC 600  
601 ACCATCAACGAGAGAGCGCGAGTGGAGCGGCTGACACCGGTGACGAGCGCGCGCGCGG 660  
601 ACCATCAACGAGAGAGCGCGAGTGGAGCGGCTGACACCGGTGACGAGCGCGCGCGCGG 660  
661 GCGCCCGGCGAGATGCGAGACCCCGCGGCGAGCATCGCGCGCGCGCACAGACACCTG 720  
661 GCGCCCGGCGAGATGCGAGACCCCGCGGCGAGCATCGCGCGCGCGCACAGACACCTG 720  
721 CAGAGCAGATGCGCTGATGACAGCAACCCCGCTGCGTGGGCGACATCTAACAAG 780  
721 CAGAGCAGATGCGCTGATGACAGCAACCCCGCTGCGTGGGCGACATCTAACAAG 780  
781 CGGTGATGATCTGTGGCGCTGAAACAAAGATCGTGGCATGTAACGCGCGGTGAGCATCTG 840  
781 CGGTGATGATCTGTGGCGCTGAAACAAAGATCGTGGCATGTAACGCGCGGTGAGCATCTG 840  
841 GACATCCGCGCAGAGCGCCCAAGAGAGCCCTTCCGCGCATAGTGAACCGCTTCTTCAAGACC 900  
841 GACATCCGCGCAGAGCGCCCAAGAGAGCCCTTCCGCGCATAGTGAACCGCTTCTTCAAGACC 900  
901 CTGCGCGCGGAGCAGGCGCAACCGAGAGCTGAAGAACTGATGACCGAGACCTGCTGTGTG 960  
901 CTGCGCGCGGAGCAGGCGCAACCGAGAGCTGAAGAACTGATGACCGAGACCTGCTGTGTG 960  
961 CAGAAACGCGCAACCCCGCATGCGAAGACCAATCTCGCGCTCTGCGCGCGCGCGCGCGCG 1020  
961 CAGAAACGCGCAACCCCGCATGCGAAGACCAATCTCGCGCTCTGCGCGCGCGCGCGCGCG 1020  
1021 GAGAGATGATGACGCGCTGCGAGGCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCTG 1080  
1021 GAGAGATGATGACGCGCTGCGAGGCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCTG 1080  
1081 GCGGAGCGATGAGCGAGCGCAACAGCGTGAACATCATGATGACAGAGAGCAATTTCAAG 1140  
1081 GCGGAGCGATGAGCGAGCGCAACAGCGTGAACATCATGATGACAGAGAGCAATTTCAAG 1140  
1141 GCGCGCGCGCGCAACGTCATGCTTCAACCTGCGCGCAAGAGAGCGCAATTTCCCAAGAC 1200  
1141 GCGCGCGCGCGCAACGTCATGCTTCAACCTGCGCGCAAGAGAGCGCAATTTCCCAAGAC 1200  
1201 TGCCGCGCGCGCGCAAGAGAGGCTGCTGAAGTGGCGGCAAGAGGCGCCACAGATGAAG 1260  
1201 TGCCGCGCGCGCGCAAGAGAGGCTGCTGAAGTGGCGGCAAGAGGCGCCACAGATGAAG 1260  
1261 GACTGACCGAGCGCGCAAGGCGCAATTTCTGGCGCAAGATCTGGCGCAACAGAGGCGCG 1320  
1261 GACTGACCGAGCGCGCAAGGCGCAATTTCTGGCGCAAGATCTGGCGCAACAGAGGCGCG 1320  
1321 CCGCGCAACTTCTCTGCAAAACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
1321 CCGCGCAACTTCTCTGCAAAACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
1381 CCGCGCGAGAGCTTCCGCTTCAAGAGAGCAACCCCGCGCGCGCGCGCGCGCGCGCGCG 1440  
1381 CCGCGCGAGAGCTTCCGCTTCAAGAGAGCAACCCCGCGCGCGCGCGCGCGCGCGCGCG 1440  
1441 CCGGAGCGCTTACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
1441 CCGGAGCGCTTACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500

QY 1501 AGCCAGTAA 1509  
Db 1501 AGCCAGTAA 1509

RESULT 3  
AX468548  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AX468548 1509 bp DNA linear PAT 16-JUL-2002  
Sequence 68 from Patent WO0226209.  
AX468548  
AX468548.1 GI:21901378  
Human immunodeficiency virus 1 (HIV-1)  
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate  
Lentivirus group.

REFERENCE  
1  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
Singh,M., Umer,J. and Dubensky,T.W.  
Microparticles for delivery of the heterologous nucleic acids  
Patent: WO 0226209-A 68 04-Apr-2002;  
CHIRON CORPORATION (US)  
Location/Qualifiers  
1. 1509  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"

ORIGIN

Query Match 99.7%; Score 1504.2; DB 6; Length 1509;  
Best Local Similarity 99.2%; Pred. No. 1e-149;  
Matches 1497; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCGCAGATCTGCGCGCGGAGAACTGGAACAGTGGAGAAATCCGC 60  
Db 1 ATGGGCGCGCGCGCGCAGATCTGCGCGCGGAGAACTGGAACAGTGGAGAAATCCGC 60  
QY 61 CTGCGCGCGCGCGCGGAGAAAGCATCACTGCTGAAGCATCTGTGTGGCGCGCGCGG 120  
Db 61 CTGCGCGCGCGCGCGGAGAAAGCATCACTGCTGAAGCATCTGTGTGGCGCGCGCGG 120  
QY 121 CTGAGAGGCTTCGCGCTTGAACCCCGCGCTGTGAGAACCGCGAGGCTGCAAGCAGATC 180  
Db 121 CTGAGAGGCTTCGCGCTTGAACCCCGCGCTGTGAGAACCGCGAGGCTGCAAGCAGATC 180  
QY 181 ATGAAGCAGCTGACAGCGCGCGCTGCGAGCCGCGACCGAGAGCTGCGAGCTGTACAAC 240  
Db 181 ATGAAGCAGCTGACAGCGCGCGCTGCGAGCCGCGACCGAGAGCTGCGAGCTGTACAAC 240  
QY 241 ACCGTGGCCACCTGTATCTGCTGTGACCGCGGATCGAGGTCCGCGACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCTGTATCTGCTGTGACCGCGGATCGAGGTCCGCGACCAAGAGGCC 300  
QY 301 CTGGAACAATGAGAGAGAGAGAAAGTCCCGAGAGAAACCCAGAGGCGCAAGAG 360  
Db 301 CTGGAACAATGAGAGAGAGAGAAAGTCCCGAGAGAAACCCAGAGGCGCAAGAG 360  
QY 361 GCGGACGCGCAAGGTGAGCGAGAACTTACCTGCTGCGAAGCTTGCAGAGGCGCGATG 420  
Db 361 GCGGACGCGCAAGGTGAGCGAGAACTTACCTGCTGCGAAGCTTGCAGAGGCGCGATG 420  
QY 421 CACCAAGGCTCAGCGCGCGCACTTGAACCGCTGGTGAAGTATGAGAGAGAGGCC 480  
Db 421 CACCAAGGCTCAGCGCGCGCACTTGAACCGCTGGTGAAGTATGAGAGAGAGGCC 480  
QY 481 TTCAAGCCCGAGAGTGAATCCCATGTTTACCGCTCTGAGGAGAGGCGCCACCCCGCAGAGC 540  
Db 481 TTCAAGCCCGAGAGTGAATCCCATGTTTACCGCTCTGAGGAGAGGCGCCACCCCGCAGAGC 540  
QY 541 CTGAACACGATGTTGAACACCGTGGGCGGCGACAGGCGCGCATGCGATGCTGAAGAGAC 600  
Db 541 CTGAACACGATGTTGAACACCGTGGGCGGCGACAGGCGCGCATGCGATGCTGAAGAGAC 600

```
QY 601 ACCATCAAGAGAGGCGCCGAGTGGAGCCGCTGACCCCGTGCAGGCGCGCCGCTG 660
|
|
|
Db 601 ACCATCAAGAGAGGCGCCGAGTGGAGCCGCTGACCCCGTGCAGGCGCGCCGCTG 660
QY 661 GCGCGCGGCGAAGTGGCGGACCCCGGCGAGGACATCGCGCGCCACAGACCTTG 720
|
|
|
Db 661 GCGCGCGGCGAAGTGGCGGACCCCGGCGAGGACATCGCGCGCCACAGACCTTG 720
QY 721 CAGGAGCAATCGCTGATGACCAAGACACCCCGCTGCGTGGGGCACTTCAAG 780
|
|
|
Db 721 CAGGAGCAATCGCTGATGACCAAGACACCCCGCTGCGTGGGGCACTTCAAG 780
QY 781 CCGTGATCATCTGCGGCTGAAACAAGATGTCGAGATGACAGCCCGTGAGATCTG 840
|
|
|
Db 781 CCGTGATCATCTGCGGCTGAAACAAGATGTCGAGATGACAGCCCGTGAGATCTG 840
QY 841 GACATCGCGCAAGGCGCCCAAGAGCCCTTCCGCACTACGTGAGCCGCTTCTTCAAG 900
|
|
|
Db 841 GACATCGCGCAAGGCGCCCAAGAGCCCTTCCGCACTACGTGAGCCGCTTCTTCAAG 900
QY 901 CTGCGCGCGGAGAGGCGCCACGAGACGTGAAGAACTGATGACCGAGACCTGCTGTG 960
|
|
|
Db 901 CTGCGCGCGGAGAGGCGCCACGAGACGTGAAGAACTGATGACCGAGACCTGCTGTG 960
QY 961 CAGAAAGCCCAAGCCGCACTGCAAGACCATCTGCGGCTCTGCGCCCGCGGCGCACTG 1020
|
|
|
Db 961 CAGAAAGCCCAAGCCGCACTGCAAGACCATCTGCGGCTCTGCGCCCGCGGCGCACTG 1020
QY 1021 GAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCGGCGCAAGGCGCGGTGCTG 1080
|
|
|
Db 1021 GAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCCCGGCGCAAGGCGCGGTGCTG 1080
QY 1081 GCGAGGCGATGAGCGGCGCAAGCGCTGAACTATGATGACGAAGAGCACTTCAAG 1140
|
|
|
Db 1081 GCGAGGCGATGAGCGGCGCAAGCGCTGAACTATGATGACGAAGAGCACTTCAAG 1140
QY 1141 GCGCGCGGCGCAAGCTCAAGTCTTCACTGCGGCAAGAGGCGCACTCGCCAAAG 1200
|
|
|
Db 1141 GCGCGCGGCGCAAGCTCAAGTCTTCACTGCGGCAAGAGGCGCACTCGCCAAAG 1200
QY 1201 TGCCTGCGCGCGCGCAAGAGGCGTGTGAAGTGGGCGAGAGGCGCAAGATGAG 1260
|
|
|
Db 1201 TGCCTGCGCGCGCGCAAGAGGCGTGTGAAGTGGGCGAGAGGCGCAAGATGAG 1260
QY 1261 GACTGACCGAGGCGCGAGGCGCACTTCCGCGGCAAGATGTTGGCCCAAGAGGCGCG 1320
|
|
|
Db 1261 GACTGACCGAGGCGCGAGGCGCACTTCCGCGGCAAGATGTTGGCCCAAGAGGCGCG 1320
QY 1321 CCGCGCAACTTCTGCAAGACCGGAGCGAGCGCGCCCGCAAGGCGCGCG 1380
|
|
|
Db 1321 CCGCGCAACTTCTGCAAGACCGGAGCGAGCGCGCCCGCAAGGCGCGCG 1380
QY 1381 CCGCGCGAGAGCTTCCGCTTCAAGAGACCAACCCCGCGCGCAAGAGCGCGAGAG 1440
|
|
|
Db 1381 CCGCGCGAGAGCTTCCGCTTCAAGAGACCAACCCCGCGCGCAAGAGCGCGAGAG 1440
QY 1441 CCGAGGCGCTTACGAGAGCGCGCTGACCGCGCTGCGAGCGCTGTTGGAGCGCGCG 1500
|
|
|
Db 1441 CCGAGGCGCTTACGAGAGCGCGCTGACCGCGCTGCGAGCGCTGTTGGAGCGCGCG 1500
QY 1501 AGCCAGTAA 1509
|
|
|
Db 1501 AGCCAGTAA 1509

RESULT 4
AX455905 1509 bp DNA linear PAT 06-JUL-2002
LOCUS Sequence 21 from Patent WO0204493.
DEFINITION AX455905
ACCESSION AX455905
VERSION AX455905.1 GI:21714897
KEYWORDS
SOURCE synthetic construct
```

```
ORGANISM synthetic construct
REFERENCE other sequences; artificial sequences.
AUTHORS 1 zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
JOURNAL Patent: WO 0204493-A 21 17-JAN-2002;
CHIRON CORPORATION (US) ; Universality of Stellenbosch (ZA)
FEATURES
source
1..1509
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="synthetic Gag coding sequence of HIV strain
AF110967"

ORIGIN
Query Match 98.7%; Score 1489.8; DB 6; Length 1509;
Best Local Similarity 99.2%; Pred. No. 3.4e-148;
Matches 1497; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGGAGAAAGCTGACAAAGTGGAGAAATCCGC 60
|
|
|
Db 1 ATGGGCGCGCGCGCGCATCTGCGCGCGGAGAAAGCTGACAAAGTGGAGAAATCCGC 60
QY 61 CTGCGCGCGCGCGGAGAGACACTACATGCTGAAGCACTGTGTGGGCGGCGCGAG 120
|
|
|
Db 61 CTGCGCGCGCGCGGAGAGACACTACATGCTGAAGCACTGTGTGGGCGGCGCGAG 120
QY 121 CTGAGAGGCTTGCCTGTAACCCCGGCTGCTGAGACCGCGAGGCTGCAACAGATC 180
|
|
|
Db 121 CTGAGAGGCTTGCCTGTAACCCCGGCTGCTGAGACCGCGAGGCTGCAACAGATC 180
QY 181 ATGAAGCACTGCAAGCCCGCTTGAAGACCGGCAACCGAGAGCTGCGGACCTGTAC 240
|
|
|
Db 181 ATGAAGCACTGCAAGCCCGCTTGAAGACCGGCAACCGAGAGCTGCGGACCTGTAC 240
QY 241 ACCGCGCGCACCTGTACTGTGTGCAAGCGCGGCAATCGAGGCGCGGACCAAGAGGCG 300
|
|
|
Db 241 ACCGCGCGCACCTGTACTGTGTGCAAGCGCGGCAATCGAGGCGCGGACCAAGAGGCG 300
QY 301 CTGACCAAGATCGAGAGAGACAGAACAAAGTCCCGAGCAAGAACCCAGAGGCCAAGAG 360
|
|
|
Db 301 CTGACCAAGATCGAGAGAGACAGAACAAAGTCCCGAGCAAGAACCCAGAGGCCAAGAG 360
QY 361 GCCGAGCGCAAGGTGAGCGAGACTACCCCATCTGTGAACCTTGCAGGCGCAAGTGTG 420
|
|
|
Db 361 GCCGAGCGCAAGGTGAGCGAGACTACCCCATCTGTGAACCTTGCAGGCGCAAGTGTG 420
QY 421 CACCAAGCGATCAAGCCCGCGGACCTGAAAGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480
|
|
|
Db 421 CACCAAGCGATCAAGCCCGCGGACCTGAAAGCGCTGGGTGAAGGTGATCGAGAGAGGCC 480
QY 481 TTCAAGCCCGAGGTGATCCCATGTTCAACCGCTGAGCGAGGCGCAACCCCGCAAGAG 540
|
|
|
Db 481 TTCAAGCCCGAGGTGATCCCATGTTCAACCGCTGAGCGAGGCGCAACCCCGCAAGAG 540
QY 541 CTGAACACAGATGTTGAACACCTGTGGGCGGCAACAGGCGCGCAAGTCTTAAGAGAC 600
|
|
|
Db 541 CTGAACACAGATGTTGAACACCTGTGGGCGGCAACAGGCGCGCAAGTCTTAAGAGAC 600
QY 601 ACCATCAAGAGAGGCGCGGAGTGGAGACCGCTGACCCCGTGCAGGCGCGCGCGCTG 660
|
|
|
Db 601 ACCATCAAGAGAGGCGCGGAGTGGAGACCGCTGACCCCGTGCAGGCGCGCGCGCTG 660
QY 661 GCGCGCGGCGAAGTGGCGGACCCCGGCGAGGACATCGCGCGCCACAGACCTTG 720
|
|
|
Db 661 GCGCGCGGCGAAGTGGCGGACCCCGGCGAGGACATCGCGCGCCACAGACCTTG 720
QY 721 CAGGAGCAATCGCTGATGACCAAGACACCCCGCTGCGTGGGGCACTTCAAG 780
|
|
|
Db 721 CAGGAGCAATCGCTGATGACCAAGACACCCCGCTGCGTGGGGCACTTCAAG 780
```

QY	761	CGGTGGATCATCTGAGGCGCTGAAACAAGATCGGCGGAGTGCAGGCCGCTGAGACATCTGTG	840
Db	761	CGCTGGATCATCTTGGGCTTGAAACAAGATCGTGGCCATGTACAGCCCCGTGAGCAATCTGTG	840
QY	841	GACATCCGCGAGGGGCCCAAGAGAGCCCTTCCGCGACTACCTAGTGAACCGCTTCTTCAAGACC	900
Db	841	GACATCCGCGAGGGGCCCAAGAGAGCCCTTCCGCGACTACCTAGTGAACCGCTTCTTCAAGACC	900
QY	901	CTGCGCGCCGAGCAGGCCACCCAGGACGTGAAGAACTGAGATGACCGAGACCTCTGTGCTG	960
Db	901	CTGCGCGCGGAGCAGGCCACCCAGGACGTGAAGAACTGAGATGACCGAGACCTCTGTGCTG	960
QY	961	CAGAAACGCAACCCCGAGCTGCAGAACCAATCCCTGGCGGCTCTCCGCGCCCGGGGCCACCTGTG	1020
Db	961	CAGAAACGCAACCCCGAGCTGCAGAACCAATCCCTGGCGGCTCTCCGCGCCCGGGGCCACCTGTG	1020
QY	1021	GAGAGATGATGACCGGCTTCGACAGGCGGTGGGCGGCGCCCGGCCACAAAGCCCGCGTGTG	1080
Db	1021	GAGAGATGATGACCGGCTTCGACAGGCGGTGGGCGGCGCCCGGCCACAAAGCCCGCGTGTG	1080
QY	1081	GCCGAGCGATGAGCCAGGCGCAACGCGTGAACATCATGATGACAGAGCAATCTTCAAG	1140
Db	1081	GCCGAGCGCATGAGCCAGGCGCAACGCGTGAACATCATGATGACAGAGCAATCTTCAAG	1140
QY	1141	GGCCCCCGGCGCAAGCTGAAGTCTTCAACTGGCGGCAAGAGGGCCACATGCGCAAGAAC	1200
Db	1141	GGCCCCCGGCGCAAGTGAAGTCTTCAACTGGCGGCAAGAGGGCCACATGCGCAAGAAC	1200
QY	1201	TGCGCGCGCCCCCGCAAGAGGGCGTGTGAAGTCCGCGCAAGAGAGGCCACCATGATGAAG	1260
Db	1201	TGCGCGCGCCCCCGCAAGAGGGCGTGTGTGAAGTCCGCGCAAGAGAGGCCACCATGATGAAG	1260
QY	1261	GACTGCACCGAGCGCCAGGCCCAATCTTCTGGGCAAGATCTGGCCAGCCACAGAGGCGCG	1320
Db	1261	GACTGCACCGAGCGCCAGGCCCAATCTTCTGGGCAAGATCTGGCCAGCCACAGAGGCGCG	1320
QY	1321	CCCGCGCACTTCTTCGCAAAACCGCAGCGAGCCCCCGGCCCCACCTGTGCCACCGCCCCC	1380
Db	1321	CCCGCGCACTTCTTCGCAAAACCGCAGCGAGCCCCCGGCCCCACCTGTGCCACCGCCCCC	1380
QY	1381	CCCGCGAGAGCTTCCGCTTGCAGGAGACCAACCCCGCGCCCGCCCAAGCAGAGGCCCAAGGAC	1440
Db	1441	CGCGAGCCCTTACCGGAGCCCTCTGACCGCCCTTGCAGACCTTGTTCGAGCGGCGCCCTGTG	1500
QY	1501	AGCCAGTAA 1509	
Db	1501	AGCCAGTAA 1509	
RESULT 5			
AX455935			
LOCUS	AX455935	1494 bp	DNA
DEFINITION	Sequence 51 from Patent WO0204493.	linear	PAT 06-JUN-2002
ACCESSION	AX455935		
VERSION	AX455935.1	GI:21714919	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	1		
TITLE	zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Renenburg, E.		
JOURNAL	Polynucleotides encoding antigenic hiv type c polypeptides,		
FEATURES	Patent: WO 0204493-A 51 17-JAN-2002;		
FEATURES	CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)		
SOURCE	1. .1494		
SOURCE	/organism="synthetic construct"		
SOURCE	/mol_type="unassigned DNA"		

Query Match	86.0%; Score 1297.6; DB 6; Length 1494;
Best Local Similarity 92.7%; Pred. No. 5,96-128;	
Matches 1402; Conservative 0; Mismatches 89; Indels 21; Gaps 3;	
QY 1 ATGAGCGCCGCGCCAGCATCTTCGCGCGCGAGAGCTGACCAAGTGGAGAGATCCGC 60	/db xref="taxon:132630"
DB 1 ATGAGCGCGCCGCGCCAGCATCTTCGAGCGCGCGCAAGCTGACCAAGTGGAGAGATCCGC 60	/note="HIV Type C Gag optimized"
QY 61 CTGGCGCCCGCGCGCAAGAGCACTAATGTGAAAGCACTGTGTGGCCAGCCGCGAG 120	
DB 61 CTGGCGCCCGCGCGCAAGAGCACTAATGTGAAAGCACTGTGTGGCCAGCCGCGAG 120	
QY 121 CTGAGAGGCTTCGCGCTTGAAACCCCGCGCTGTGTGAGACCCGCGAGGGCTGCAACATC 180	
DB 121 CTGAGAGGCTTCGCGCTTGAAACCCCGCGCTGTGTGAGACCCAGCGGGCTGCAACATC 180	
QY 181 ATGAGAGAGCTGACGCCCGCTTGACAGCCGCGCACCGAGAGCTGGCGACCTGTACAC 240	
DB 181 ATGAGAGAGCTGACGCCCGCTTGACAGCCGCGCACCGAGAGCTGGCGACCTGTACAC 240	
QY 241 ACCGTGGCGACCCCTGTACTGTGTGTGCAAGCGCGGACTGAGGTCCGCGACCAAGAGAGCC 300	
DB 241 ACCGTGGCGACCCCTGTACTGTGTGTGCAAGAGGACTGAGGTGTGCGACCAAGAGAGCC 300	
QY 301 CTGGAACAAGATCGAGAGAGAGCAGAACAAAGTCCGAGCAGAAACCCAGCAGGCGCAAGAG 360	
DB 301 CTGGAACAAGATCGAGAGAGAGCAGAACAAAGTCCGAGCAGAAAGGCCAGAGCCAGAGCC 360	
QY 361 GCCAGCGGCAAGGTGAGCCAGAACTACCCACTGTGTGAAACTGTGCAAGGCGCAGATGTG 420	
DB 361 GCCAGCGGCAAGGTGAGCCAGAACTACCCACTGTGTGAAAGCCAGAGGCGCAGATGTG 420	
QY 421 CACCAAGGCGCATCAGACCCCGCGACCCCTTGACCGCTGTGGTGAAGGTGATTCAGAGAAAGGCC 480	
DB 421 CACCAAGGCGCATCAGACCCCGCGACCCCTTGACCGCTGTGAATCAAGGTGATTCAGAGAAAGGCC 480	
QY 481 TTCAAGCCCGAGGTGATTCGCCATTTTCAACGCGCTTGAGCGAGGCGCACCCCGCAGAG 540	
DB 481 TTCAAGCCCGAGGTGATTCGCCATTTTCAACGCGCTTGAGCGAGGCGCACCCCGCAGAG 540	
QY 541 CTGAACAAGATGTGAAACACCGTGGGCGCGCACCGAGCCGCGCATGACATGCTGAAGAGAC 600	
DB 541 CTGAACAAGATGTGAAACACCGTGGGCGCGCACCGAGCCGCGCATGACATGCTGAAGAGAC 600	
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGACCGCTGCAACCCCGTGCAGAGCGCGGCCGTG 660	
DB 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGACCGCACCCCGTGCAGAGCGCGGCCGTG 660	
QY 661 GCCCGCGCGAGAGAGCGCGACCCCGCGGAGCGACATTCGCGCGCGCACAGACCCCTG 720	
DB 661 GCCCGCGCGAGAGAGCGCGACCCCGCGGAGCGACATTCGCGCGCGCACAGACCCCTG 720	
QY 721 CAGGAGAGAGATCGCGCTGATGACAGAGAACCCCGCTGCGCGCTGTGGCGCATCTTAAG 780	
DB 721 CAGGAGAGAGATCGCGCTGATGACAGAGAACCCCGCATTCGCGCGAGACATCTTAAG 780	
QY 781 CGGTGATCATCTTGCGCTTGAAACAGATGTGTGATGACAGCCCGCTGTAGCATCTTG 840	
DB 781 CGGTGATCATCTTGCGCTTGAAACAGATGTGTGATGACAGCCCGCTGTAGCATCTTG 840	
QY 841 GACATCGGCAAGGCGCGCACAGAGCCCTTCGCGGACATAGTGAGACGCGCTTCTTAAGACC 900	
DB 841 GACATCGGCAAGGCGCGCACAGAGCCCTTCGCGGACATAGTGAGACGCGCTTCTTAAGACC 900	
QY 901 CTGGCGCGCGAGCAGGCGCACCCAGAGAGTGAAGAACTGTATGACCCGAGACCCCTGTGTG 960	
DB 901 CTGGCGCGCGAGCAGGCGCACCCAGAGAGTGAAGAACTGTATGACCCGAGACCCCTGTGTG 960	
QY 961 CAGAGCGCAACCCCGACTGCAAGACCATCTTGCGCGCTTCGAGCCCGGCGCACCTGT 1020	

[illegible]

Db	61	CTGCGCCCGGCGCGCGCAGCACTACATGCTGAAGCACTGGTGTGGCCAGCCGCGAG	120
Qy	121	CTGAGAGGGCTTCGCGCTTGAACTCCCGGCGCTGTGAAGACCGCGAGGGCTGCAAGCAAGT	180
Db	121	CTGAGAGCGCTTCGCGGTGAACCCCGGCGCTGTGAAGACCGAGGACGGCTGGCAGATC	180
Qy	181	ATGAAGCAGCTGACACCCCGCTGTGAAGACCGGACACCGAGAGCTGGACCTGTACAA	240
Db	181	ATCAAGCAGCTGACACCCCGCTGTGAAGACCGGACACCGAGAGATCCGACGCTGTTCAC	240
Qy	241	ACCGTGGCCACCTCTGATCTGCTGTGCAACCGCGCATGAGGTCCGAGACCAAGAGAGCC	300
Db	241	ACCGTGGCCACCTCTGATCTGCGGTGCAAAAGGGCATGACGTGGCGGACACCAAGAGAGCC	300
Qy	301	CTGGAACAAGTTCGAGAGAGAGAGCAAAAGTCCGACGAGAAAGCCACAGGCCCAAGAG	360
Db	301	CTGGAACAAGATTCGAAGAGAGAGCAAAAGTCCGACGAGAAAGCCACAGGCCCAAGAGCC	360
Qy	361	GCCGACGGCAAGGTGAGCGCAGAACTAACCCCATCGTGCAAACTTGCAGAGGCCAGATGTG	420
Db	361	GCCGACAAAGAGGTGAGCGCAGAACTAACCCCATCTGTGCAAACTTGCAGAGGCCAGATGTG	420
Qy	421	CACCAAGCCATCAGCCCCCGCACCTCTGAAGCTTGGTGAAGGTGATGAGAGAAAGGCC	480
Db	421	CACCAAGCCATCAGCCCCCGCACCTCTGAAGCTTGGTGAAGGTGATGAGAGAAAGGCC	480
Qy	481	TTGACGCCCCGAGGTGATGCCCATGTTCACCGGCGTGAGCGAGGGGGCCACCCCGCAGAC	540
Db	481	TTGACGCCCCGAGGTGATCCCATGTTCACCGGCGTGAGCGAGGGGGCCACCCCGCAGAC	540
Qy	541	CTGAACAAGATGTTGAACACCGTGGGCGGCACAGGCGGCATGACAGATCTGAAGAGC	600
Db	541	CTGAACAAGATGTTGAACACCGTGGGCGGCACAGGCGGCATGACAGATCTGAAGAGC	600
Qy	601	ACCATCAACGAGAGAGCGCCGCGAGTGGGACCCGCTTGACCCCGTGTGAGGCGGCGCCGTG	660
Db	601	ACCATCAACGAGAGAGCGCCGCGAGTGGGACCCGCTTGACCCCGTGTGAGGCGGCGCCGTG	660
Qy	661	GCCCCCGGCGCAGATGCGGCGAACCCCGCGGCGACGACATCGCGGCGCCACAGCACCTG	720
Db	661	GCCCCCGGCGCAGATGCGGCGAACCCCGCGGCGACGACATCGCGGCGCCACAGCACCTG	720
Qy	721	CAGAGCAGATGCGCTGTGATACCAAGCACTCCCGCTGTGCGGTGGGCGACATCTCAAG	780
Db	721	CAGAGCAGATGCGCTGTGATACCAAGCACTCCCGCTGTGCGGTGGGCGACATCTCAAG	780
Qy	781	CGGTGATCATCTGGGGCTGGAACAAGATCGTGGAGATGTACAGCCCGTAGAGATCTGT	840
Db	781	CGGTGATCATCTGGGGCTGGAACAAGATCGTGGAGATGTACAGCCCGGTGAGATCTGT	840
Qy	841	GACATCCGCGCAGGGCCCCCAAGAGACCCCTTCGCGACATACGTGAACGCGCTTCTTCAAGCC	900
Db	841	GACATCAAGCAGGGCCCCCAAGAGACCCCTTCGCGACATACGTGAACGCGCTTCTTCAAGCC	900
Qy	901	CTGGCGCGCGAGCGACCAACCAAGAGGTGAAGAACTGAGTGAACCACTCTGTGTGTG	960
Db	901	CTGGCGCGCGAGCGACCAACCAAGAGGTGAAGAACTGAGTGAACCACTCTGTGTGTG	960
Qy	961	CAGAACGCAACCCCGACTGCAAGACCATCTGTGCGGCTCTCGGCGCGCGGCGCACCTGT	1020
Db	961	CAGAACGCAACCCCGACTGCAAGACCATCTGTGCGGCGCGGCGCGCGGCGCACGCTGT	1020
Qy	1021	GAGGAGATGATGACCGGCTGTCAAGGGGTGGGCGGCGCCCGGCAAGAGCCCGGTGTGTG	1080
Db	1021	GAGGAGATGATGACCGGCTGTCAAGGGGTGGGCGGCGCCCGGCAAGAGCCCGGTGTGTG	1080
Qy	1081	GCCGAGCGATGAGCCAGGCGCAACAGCGGTGAACATCACTGATGTCGAAGAGAGCACTTCAAG	1140
Db	1081	GCCGAGCGCATGAGCCAGGCGCAACCAAGCGGTGATGATCAGAAAGAGCACTTCAAG	1140
Qy	1141	GCGCCCGCGCAAGTCAAGTCTTCAACTGCGGCAAGAGGGCGCATTCGCCAAGAAC	1200

Db 1141 GGCCCCCGGCGCGGTGAAGTCTTCAACTGCGCGCGGAGGCCACATCCGCCGCAAC 1200  
Qy 1201 TCCCGCGCCCCCGCAAGAGGCTCTGTGAAGTGGCGCAGAGAGGCCACCATGATGAAG 1260  
Db 1201 TCCCGCGCCCCCGCAAGAGGCTCTGTGAAGTGGCGCAGAGAGGCCACCATGATGAAG 1260  
Qy 1261 GACTGACCGGAGCGGCGCAACTCTCTGGGCAAGATCTGGGCGGCGGCAAGAGGCGCG 1320  
Db 1261 GACTGACCGGAGCGGCGCAACTCTCTGGGCAAGATCTGGGCGGCGGCAAGAGGCGCG 1320  
Qy 1321 CCGGCAACTCTCTGTGAAGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1377  
Db 1321 CCGGCAACTCTCTGTGAAGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
Qy 1378 CCCCCCGCGAGAGCTTCCGCTTTCAGAGAGACACCCCGCGCGGCGGCGGCGGCGGCGG 1437  
Db 1381 CCCCCCGCGAGAGCTTCAAGTTCAGAGAGACACCCCGCGGCGGCGGCGGCGGCGGCGG 1431  
Qy 1438 GACCGGAGCCCTTACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1497  
Db 1432 G-----ACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1479  
Qy 1498 CTGAGCCAGTAA 1509  
Db 1480 CTGAGCCAGTAA 1491

RESULT 7  
LOCUS AY181195 1485 bp DNA linear SYN 02-OCT-2003  
DEFINITION Synthetic construct HIV-1-derived gag protein (gag) gene, complete  
ACCESSION AY181195  
VERSION AY181195.1 GI:37413992  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
synthetic construct  
other sequences; artificial sequences.

REFERENCE  
AUTHORS  
1 (bases 1 to 1485)  
Gao, F., Li, Y., Decker, J. M., Peyerl, F. W., Bibollet-Ruche, F.,  
Rodenburg, C. M., Chen, Y., Shaw, D. R., Allen, S., Musonda, R.,  
Shaw, G. M., Ziaja, A. J., Letvin, N. and Hahn, B. H.  
Codon usage optimization of HIV type 1 subtype C gag, pol, env, and  
nef genes: in vitro expression and immune responses in  
DNA-vaccinated mice  
AIDS Res. Hum. Retroviruses 19 (9), 817-823 (2003)  
2 (bases 1 to 1485)  
Gao, F., Li, Y., Decker, J. M., Bibollet-Ruche, F., Rodenburg, C. M.,  
Chen, Y., Shaw, D. R., Allen, S., Musonda, R., Ziaja, A. J. and  
Hahn, B. H.  
Direct Submision  
Submitted (16-NOV-2002) Medicine, Duke University Medical Center,  
112 Rptll, Lasalle Street, Durham, NC 27710, USA

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Hahn, B. H.  
Direct Submision  
Submitted (16-NOV-2002) Medicine, Duke University Medical Center,  
112 Rptll, Lasalle Street, Durham, NC 27710, USA

FEATURES  
SOURCE  
1..1485  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="codon usage optimization of HIV-1 96ZM651"  
1..1485  
/gene="gag"  
1..1485  
/gene="gag"  
/codon\_start=1  
/transl\_table=11  
/product="gag protein"  
/protein\_id="AA065393.1"  
/db\_xref="GI:37413993"  
/translation="MGARSTIRRGGLDKMEKIRLPQKKRIMIKHLVMSRELIERF  
ALNPLLETSECKQIMKQLOPALQTEELSLVNTVAATLVCHEGVAVRTKALD  
RIEBOONIQOKIIOKTOADGKVSQNPVONLQGVHOKLSPTLMAVKTIEE  
KAFSEVIMPTALSEGATPDNTMLNTVGGHQAQMOKLQTIINEBAEMRLHPVH  
AGPIAPGQMRPRGSDIAGTSTTLEQJAMTMSNPPIPGDIYKRMIIIGLKIIVMY

gene  
CDS

SPVSLTIDIKQEPKPRDYVDFPFKTLBAEQTOBVKMMTDTLLVONANPDCKTLK  
ALGPRATLEBMMTACQGVGSHKARVALAEMASQNSVILIMKSNFGNKMVCFN  
CKEKSHITANRCAPRKPKGCKWKCKGKHQMKCTEONANPLGKINSHKGRPQNPLONR  
PEPTAPPAESFBEETTPAPKQESKOREALTSLSKLSFSDPLSO"

ORIGIN  
Query Match 85.6%; Score 1291.4; DB 12; Length 1485;  
Best Local Similarity 92.6%; Pred. No. 2.7e-127;  
Matches 1403; Conservative 0; Mismatches 76; Indels 36; Gaps 3;

Qy 1 ATGGGCGCGCGCGCGCGCATCTCTGCGCGGCGGAGAGCTGAGCAAGTGGAGAAATCCGC 60  
Db 1 ATGGGCGCGCGCGCGCGCATCTCTGCGCGGCGGAGAGCTGAGCAAGTGGAGAAATCCGC 60  
Qy 61 CTGGCGCGCGCGCGCGCGCATCTCTGAGAGCACCTGTGTGTGGGCGAGCGCGGAG 120  
Db 61 CTGGCGCGCGCGCGCGCGCATCTCTGAGAGCACCTGTGTGTGGGCGAGCGCGGAG 120  
Qy 121 CTGGAGGGGCTTGCGCCTGAAACCGGCGCTGTGAGAGACCGGAGGGCTGCAAGCAGATC 180  
Db 121 CTGGAGGGGCTTGCGCCTGAAACCGGCGCTGTGAGAGACCGGAGGGCTGCAAGCAGATC 180  
Qy 181 ATGAAGCAGCTGCGACCGCGCCTGTGAGACCGGCGACCGAGAGCTGCGACGCTGTACAA 240  
Db 181 ATGAAGCAGCTGCGACCGCGCCTGTGAGACCGGCGACCGAGAGCTGCGACGCTGTACAA 240  
Qy 241 ACCGTGGCCACCTGTGATCTGCGTGCACCGCGGCGATCGAGTCTCGGACACCAAGAGGCC 300  
Db 241 ACCGTGGCCACCTGTGATCTGCGTGCACCGCGGCGATCGAGTCTCGGACACCAAGAGGCC 300  
Qy 301 CTGGACCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 355  
Db 301 CTGGACCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 355  
Qy 356 -AGGAGCGCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 414  
Db 356 -AGGAGCGCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 414  
Qy 415 ATGTGTCACACAGAGCATAGCGCCCGCACCTGTGAACGCTGTGGTGAAGTATTCAGAGAG 474  
Db 415 ATGTGTCACACAGAGCATAGCGCCCGCACCTGTGAACGCTGTGGTGAAGTATTCAGAGAG 474  
Qy 475 AAGGCTTCAGCGCGGAGAGTATCCCAATGTTACCGGCGCTGAGGAGAGGCGCGCACCGCC 534  
Db 475 AAGGCTTCAGCGCGGAGAGTATCCCAATGTTACCGGCGCTGAGGAGAGGCGCGCACCGCC 534  
Qy 535 CAGGACTGAAACCGATGTTGAACAACCGTGGCGGCGCACAGGCGCGCATGAGATGCTG 594  
Db 535 CAGGACTGAAACCGATGTTGAACAACCGTGGCGGCGCACAGGCGCGCATGAGATGCTG 594  
Qy 595 AAGGACACCATCAACGAGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 654  
Db 595 AAGGACACCATCAACGAGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 654  
Qy 655 CCGGAGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 714  
Db 655 CCGGAGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 714  
Qy 715 ACCCTGAGAGAGAGATGCGCTGTGATGACAGCAACCGCCCGTGTGCGGCGAGCATC 774  
Db 715 ACCCTGAGAGAGAGATGCGCTGTGATGACAGCAACCGCCCGTGTGCGGCGAGCATC 774  
Qy 775 TACAAAGCGGTGATCATCTGTGGGCTGTGAACAAGATCTGTGGATGTACAGCCCGTGTAGC 834  
Db 775 TACAAAGCGGTGATCATCTGTGGGCTGTGAACAAGATCTGTGGATGTACAGCCCGTGTAGC 834  
Qy 835 ATCCGAGCATCCGCGAGAGGCGCGCAAGAGAGCGCTTCGCGACTAGTGGAGCGCTTCTTC 894  
Db 835 ATCCGAGCATCCGCGAGAGGCGCGCAAGAGAGCGCTTCGCGACTAGTGGAGCGCTTCTTC 894  
Qy 895 AAGACCTGTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954  
Db 895 AAGACCTGTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 954  
Qy 901 AAGACCTGTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 901 AAGACCTGTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

Qy		955	CTGGTGCAGAAACCCCAACCCCCGATCTGCAGAACATTCTGCGGTCTTG36CCC	CGGCGCC	1014
Dd		961	CTGGTGCAGAAACCCCAACCCCCGATCTGCAGAACATTCTGCAGGCCCTTG36CCC	CGGCGCC	1020
Qy		1015	ACCCTGAAGAATGATGACCGCTTCGACGGGGGTGGGGGGGGCCCCGGCCACAAGGCCGC		1074
Dd		1021	ACCCTGAAGAATGATGACCGCTTCGACGGGGGTGGGGGGGGCCCCGGCCACAAGGCCGC		1080
Qy		1075	GTCCTGGCCGAGGCGCATGAGCCAGGCCAAACAGCTGTAACATCATGATGACGAAGAAGCAAC		1134
Dd		1081	GTCCTGGCCGAGGCGCATGAGCCAGGCCAAACAGGTTGAATCTCTATGTGAAGAAGCAAC		1140
Qy		1135	TTCGAAGGGCCCCCGGGGCAACTGACAGTCTTCAAATGCGGCAAGAGAGGCCAATCGCC		1194
Dd		1141	TTCAAGGGCAACAGAGGCATGTGTGAAGTGTCTTCAACTGCGGCMAAGAGAGGCCAATCGCC		1200
Qy		1195	AAGAACTGCGCGCCCGCCCGCAAGAAAGGGCTGTG9AAGTGGAGGAAGAGGGCCACAG		1254
Dd		1201	CGCACTGCGCGCGCCCGCCCGCAAGAGGGCTGTG9AAGTGGAGGAAGAGGGCCACAG		1260
Qy		1255	ATGAAGACTGCAACCGAGGCGCAGGCCAACTTCTG9GCAAGATCTG9CCAGCCAAG		1314
Dd		1261	ATGAAGACTGCAACCGAGGCGCAGGCCAACTTCTG9GCAAGATCTG9CCAGCCAAG		1320
Qy		1315	GGCGCGCGCGGACCTTCTGTGAGAAACCGCACGAGCCCGCGGCCCAACCTGTGCCAC		1374
Dd		1321	GGCGCGCGCGGACCTTCTGTGAGAAACCG-----CCCGAGGCCAC		1362
Qy		1375	GGCGCGCGCGCGAGAGCTTCCGCTTCGAGAGAACACCCCGCGCCCAAGAGAGCC		1434
Dd		1363	GGCGCGCGCGCGAGAGCTTCCGCTTCGAGAGAACACCCCGCGCCCAAGAGAGAGC		1422
Qy		1435	AAGAACCGGAGACCTTACCGGAGCCCTGTACCGCCTTGGCGAGCTTGTGGAGCGGC		1494
Dd		1423	AAGG-----ACCGGAGCGCCTGTACCMAGCTTGAAGAGCTTGTGGAGCGAGC		1470
Qy		1495	CCCTGAGCCAGTAA	1509	
Dd		1471	CCCTGAGCCAGTAA	1485	
<hr/>					
RESULT 8					
AX455887					
LOCUS	AX455887	1479 bp	DNA	linear	PAT 06-JUL-2002
DEFINITION	Sequence 3 from Patent WO0204493.				
ACCESSION	AX455887				
VERSION	AX455887.1 GI:21714880				
KEYWORDS	.				
SOURCE					
ORGANISM	synthetic construct				
REFERENCE	synthetic construct				
AUTHORS	other sequences; artificial sequences.				
TITLE	zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.				
JOURNAL	Polynucleotides encoding antigenic hiv type c polypeptides,				
FEATURES	Patent: WO 0204493-A 3 17-JAN-2002;				
source	CHIRON CORPORATION (US); University of Stellenbosch (ZA)				
	Location/Qualifiers				
	1..1479				
	/organism="synthetic construct"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:32630"				
	/note="Synthetic Gag of HIV strain AF110965"				
<hr/>					
ORIGIN					
Query Match	84.6%; Score 1276.8; DB 6; Length 1479;				
Best Local Similarity	92.5%; Pred. No. 9.2e-126;				
Matches 1399; Conservative	0; Mismatches 77; Indels 36; Gaps 4;				
Qy		1	ATGGGCGCGCGCGCAGATCTCTGCGGGCGAGAAAGCTGAGCAAGTGGAGAAAGATCCG		60
Dd		1	ATGGGCGCGCGCGCAGATCTCTGCGGGCGAGCGGCAAGCTGAGACGCTTGGAGGCCATCCG		60

QY	61	CTGCGCCCGCGCGCAAGAGCATATCATGCTGAAGCACTGATGTGGCCAGCCGCGAG	120
Db	61	CTGCGCCCGCGCGCAAGAGGTCTATCATGTAGAGCACTTGATGTGGCCAGCCGCGAG	120
QY	121	CTGAGAGGCTTGGCCCTGTAACCCCGGCTGTCTGAGAACCGCGAGGGCTGCAAGCATTC	180
Db	121	CTGAGAAATTGGCCCTGTAACCCCGGCTGTCTGAGAACCGAGGAGGGCTGCAAGCATTC	180
QY	181	ATGAAGCAGCTCAGCGCCGCTGAGACCGGACACGAGAGACTGGCGAGCTGTATCAAC	240
Db	181	ATCCCGCAGCTCAGCCCGGCTGTGAGACCGGACACGAGAGACTGAAGGCTGTTCAC	240
QY	241	ACCGTGGCCACCTCTGATCTGCTGCACGCGGCACTGAGGTCCGCGACCAAGAGGCTC	300
Db	241	ACCGTGGCCACCTCTGATCTGCTGCACGAGAGTCTGAGGTCCGCGACCAAGAGGCTC	300
QY	301	CTGGAACAAGATGAGAGAGGAGCAACAAGTCCGACGAGAAACCAAGAGGCGCAAGAG	360
Db	301	CTGGAACAAGATGAGAGAGGAGCAACAAGTCCGACGAGAAATCCAGAGGCGCGAGGCTC	360
QY	361	GGCGA-C--GGCAAGGTGAGCCAGAACTAACCCATCTGTGCAAACTGTGAGGCGCAGATG	417
Db	361	GGCGAACAAGGCAAGGTGAGCCAGAACTAACCCATCTGTGCAAACTGTGAGGCGCAGATG	420
QY	418	GTGACACAGGCGCATGAGCCCGCGACCCCTGAAACGCTGGGTGAAAGTGTGAGAGAAAG	477
Db	421	GTGACACAGGCGCATGAGCCCGCGACCCCTGAAACGCTGGGTGAAAGTGTGAGAGAAAG	480
QY	478	GGCTTCAAGCCCGAGGTATCCCATGTTCTAACCGGCTGTAGCGAGGGCGCACCCCGAG	537
Db	481	GGCTTCAAGCCCGAGGTATCCCATGTTCTAACCGGCTGTAGCGAGGGCGCACCCCGAG	540
QY	538	GACCTGAAACAGATTTTGAACAACGTGGGGCGGCGACAGGGCGGCAATGCAATGCTGAAG	597
Db	541	GACCTGAAACAGATTTTGAACAACGTGGGGCGGCGACAGGGCGGCAATGCTGAAG	600
QY	598	GACACATCAACAGAGAGGCGCGAGGTGGAGCGGCTGCAACCCCGTCAAGGCGGCGCC	657
Db	601	GACACATCAACAGAGAGGCGCGAGGTGGAGCCCGGTGCAACCCCGTCAAGGCGGCGCC	660
QY	658	GTGGCCCCCGGCGAGATCGCGAACCCCGCGGACGCACTATCGCGGCGCACAGCAC	717
Db	661	ATCGCCCCCGGCGAGATCGCGAGGCCCCGCGGACGCACTATCGCGGCGCACAGCAC	720
QY	718	CTGCAAGAGCAGATGACTGTGATGACAGCAACCCCGCGGCGGTGGGAGCATCTAC	777
Db	721	CTGCAAGAGCAGATGACTGTGATGACAGCAACCCCGCGGCGGTGGGAGCATCTAC	780
QY	778	AAGCGGTGATCATCTTGAGGCTTGAACAAGATCGTGCGATGTACAGGCCGCTGAGCATC	837
Db	781	AAGCGGTGATCATCTTGAGGCTTGAACAAGATCGTGCGATGTACAGGCCGCTGAGCATC	840
QY	838	CTGCAACATCCGCGCAGGCGCCCAAGAGCCCTTCCGCGCATCTGTGAGACCGGTTCTTCAAG	897
Db	841	CTGCAACATCAAGCAGGCGCCCAAGAGCCCTTCCGCGCATCTGTGAGACCGGTTCTTCAAG	900
QY	898	AACCTGCGCGCGAGCAGGCGCACCCAGGACGTGAAGAACTGAGATGACCGAAGCCTGTGTG	957
Db	901	AACCTGCGCGCGCGAGCAGGCGCACCCAGGAGGTGAAGAACTGAGATGACCGAAGCCTGTGTG	960
QY	958	GTGCAAAACCGCAACCCCGATGCAAGACATCTCTGGCGCTTCTGGGCCCGGCGCCAC	1017
Db	961	GTGCAAAACCGCAACCCCGATGCAAGACATCTCTGGCGCTTCTGGGCCCGGCGCCAC	1020
QY	1018	CTGAGAGAGATGATGACCGCTGCGCAGGCGGTGGGCGGCGCCCGGCGCAAGAGCCGCGGTG	1077
Db	1021	CTGAGAGAGATGATGACCGCTGCGCAGGCGGTGGGCGGCGCCCGGCGCAAGAGCCGCGGTG	1080
QY	1078	CTGGCCGAGGCGATGAGCGCAGGCGCAACGCGTGAACATCATGATGCAAGAGCACTTC	1137
Db	1081	CTGGCCGAGGCGATGAGCGCAGGCGCAACGCGTGAACATCATGATGCAAGAGCACTTC	1137

QY	1138	AAGGGCCCCGGCGCAAGTCAAGGCTTCAACTGCGGCACAGAGGGCCATATGCCAAG	1197
Db	1138	AAGGGCCCCGGCGGATGTCTAAGTGTCTTCAACTGCGGCAGAGAGGCCACATATGCCCGC	1197
QY	1198	AACCTCCGCGCCCCCGCAGAGAGGGCTGTGGAAGTCGGCAAGAGAGGCCACCAAGATG	1257
Db	1198	AACTGCGCGCCCCCGCAGAGAGGGCTGTGGAAGTCGGCAAGAGAGGCCACCAAGATG	1257
QY	1258	AAGACTGCACCGAGCGCCAGGGCCAACTTCTGGGCAAGATCTGGCCCAAGCACAAGGGC	1317
Db	1258	AAGACTGCACCGAGCGCCAGGGCCAACTTCTGGGCAAGATCTGGCCCAAGCACAAGGGC	1317
QY	1318	CGCCCCGGGCACTTCTCTGCAGAACCGCGACGACGCGCGCGCCCAACCTGTCCACCGCC	1377
Db	1318	CGCCCCGGGCACTTCTCTGCAGAGCGG-----CCCGAGCCCAACCGCC	1359
QY	1378	CCCCCGCGAGAGCGCTTCGCGCTTCGAGGAGACACCCCCCGGCCCAAGAGAGGCCCAAG	1437
Db	1360	CCCCCGCGAGAGCTTCCGCTTCGAGGAGACACCCCCCGGCCAAGAACAGAGAGGACAG	1419
QY	1438	GACCGGAGCCCTACCGCGAGCGCCCTGTACCGCGCTTCGCGACACTTTCGCGAGCGGCGCC	1497
Db	1420	G-----ACCGGAGACCTGTACCGAGCCTGAAGAGCGCTTTCGCGACAGACCC	1467
QY	1498	CTGAGCCAGTAA	1509
Db	1468	CTGAGCCAGTAA	1479

RESULT 9				
LOCUS	AX468543			
DEFINITION	AX468543 Sequence 63 from Patent WO026209.	1479 bp	DNA	
ACCESSION	AX468543			linear
VERSION	AX468543.1	GI:21901373		PAT 16-JUL-2002
KEYWORDS				
SOURCE				
ORGANISM	Human immunodeficiency virus 1 (HIV-1)			
	Human immunodeficiency virus 1			

REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	SOURCE
1	O'Hagan, D., Otten, G., Donnelly, J. J., Polo, J. M., Barnett, S., Singh, M., Ulmer, J., and Dubensky, T. W.	Microparticles for delivery of the heterologous nucleic acids	Parent. WO 0226209 A 63 04-Apr-2002; CHIRON CORPORATION (US)	Location/Qualifiers	1. 1479 /organism:"Human immunodeficiency virus 1" /mol_type:"unassigned DNA" /db_xref="taxon:11676"

**ORIGIN**

Query Match	84.6%	Score 1276.8;	DB 6;	Length 1479;
Best Local Similarity	92.5%	Pred. No. 9.2e-126;		
Matches 1399;	Conservative 0;	Mismatches 77;	Indels 36;	Gaps 4;

Qy	1	ATGGGCGCCCGGCGCAGCATCTCTGGCGGGCGAGAGCTGTGA/CA/ATGTGGAGAAAGATCCGC	60
Db	1	ATGGGCGCCCGGCGCAGCATCTCTGGCGGGCGAGAGCTGTGA/CGCCCTGGAGAGCGCATCCCC	60
Qy	61	CTGGCGCCCGGCGGCGAGAGAGCATCAATGCTGAAGCA/CTGGTGTGGGCGAGCGCGAG	120
Db	61	CTGGCGCCCGGCGGCGAGAGAGTGTCAATGATAGAGCA/CTGGTGTGGGCGAGCGCGAG	120
Qy	121	CTGGAGGGGCTTTCGCCCTTGAA/CCCGCGGCTGTGAGAACCGCGAGGGCTGCAAGCAGATC	180
Db	121	CTGGAGAGTTTCGCCCTTGAA/CCCGCGGCTGTGAGAACCAAGCGAGGGCTGCAAGCAGATC	180
Qy	181	ATGAAGCAGCTTCAGACCGCGCCCTTGAGAACCGGCA/CCGAGAGACTGTGCGAGCCTGTTCAC	240
Db	181	ATTCGCGCAGCTCAGACCGCGCCCTTGAGAACCGGCA/CCGAGAGACTGTAAAGGCTGTTCAC	240

QY	241	ACCGGCGCAACCCCTTA	CTGGCTGCAACCCCGGCATCTGAGGTC	CCGGGACACCAAGAGGC	300	
Db	241	ACCGGCGCACCCCTGTACTGGCTGCA	CGAAGAA	GATCTGAGGTC	CCGGACACCAAGAGGC	300
QY	301	CTGGCAAGATCTGAGAGAGAGCA	CAAGTCC	CCAGCAAGAACCAAGAGGC	CAAGAG	360
Db	301	CTGGCAAGATCTGAGAGAGAGCA	CAAGTCC	CCAGCAAGAACCAAGAGGC	CAAGAG	360
QY	361	GCCTGCAAGAGGCAAGGTGAGGCA	GAATACCC	CATCTGTCAGAA	CTTGCAAGGCGCAGATG	420
Db	361	GCCTGCAAGAGGCAAGGTGAGGCA	GAATACCC	CATCTGTCAGAA	CTTGCAAGGCGCAGATG	420
QY	418	GTGCAACAGGCGCATAGCC	CCCCCGGCACCTCTGAACCGCTGGGTGA	AACTGATCGAAGAGAG	477	
Db	421	GTGCAACAGGCGCATAGCC	CCCCCGGCACCTCTGAACCGCTGGGTGA	AACTGATCGAAGAGAG	480	
QY	478	GCCTTCAACCCCGAGGTATCC	CCATGTTCAACCGCCCTGAGGAGAGGCGCCAC	CCCCCGAG	537	
Db	481	GCCTTCAACCCCGAGGTATCC	CCATGTTCAACCGCCCTGAGGAGAGGCGCCAC	CCCCCGAG	540	
QY	538	GACCTGAACAGATGTTGA	CAACCGTGGGCGGCCACAGAGCGGCATGCA	ATGCTGAAG	597	
Db	541	GACCTGAACAGATGTTGA	CAACCGTGGGCGGCCACAGAGCGGCATGCA	ATGCTGAAG	600	
QY	558	GACACATCAACAGAGAGCGCGCGAG	GTGGACCGCCTGCAACCCCGTGCAGGCGCGCC	657		
Db	601	GACACATCAACAGAGAGCGCGCGAG	GTGGACCGCCTGCAACCCCGTGCAGGCGCGCC	660		
QY	658	GTGGCGCCCGGCGAGATGCG	CGACCCCGCGCGAGCGACATCGCGCGGCGCCACAGACCC	717		
Db	661	ATCGCCCCCGGCGAGATGCG	CGACCCCGCGCGAGCGACATCGCGCGGCGCCACAGACCC	720		
QY	718	CTGCAGAGAGAGATGCTG	CTGTGATGACAGCAACCC	CCCCCATCCCGTGGGCGACATCTAC	777	
Db	721	CTGCAGAGAGAGATGCTG	CTGTGATGATGACAGCAACCC	CCCCCATCCCGTGGGCGACATCTAC	780	
QY	778	AAAGCGTGGATCATCTTG	GGCCTTGAA	CAAGATCGTGGAGATGACAGCCCGGTGAGCATC	837	
Db	781	AAAGCGTGGATCATCTTG	GGCCTTGAA	CAAGATCGTGGAGATGACAGCCCGGTGAGCATC	840	
QY	838	CTGCAACATCCCGCAGGGG	CCCCCAAGAGGCGCTTCGCGACATACGTGGACCGCTTCTTCAAG	897		
Db	841	CTGCAACATCAAGAGGGG	CCCCCAAGAGGCGCTTCGCGACATACGTGGACCGCTTCTTCAAG	900		
QY	898	AACCTGCGCGCGAGCAGG	CCACCCAGGACGTGAAGAACTG	AGATGACAGAGCCCTGCTG	957	
Db	901	AACCTGCGCGCGAGCAGG	CCACCCAGGAGTGAAGAACTG	AGATGACAGACCCCTGCTG	960	
QY	958	GTGCAGAACGCGCAACCC	CGACTGCAAGACATCTCGCGGCTCTTG	GGCCCCGCGCGCAC	1017	
Db	961	GTGCAGAACGCGCAACCC	CGACTGCAAGACATCTCGCGGCTCTTG	GGCCCCGCGCGCAC	1022	
QY	1018	CTGAGAGAGATGATGACCG	CTCTGCAAGGGCGTGGGCGGCCCCCGGCA	CAAGGCGCGCTG	1077	
Db	1021	CTGAGAGAGATGATGACCG	CTCTGCAAGGGCGTGGGCGGCCCCCGGCA	CAAGGCGCGCTG	1080	
QY	1078	CTGGCCGAGCGATGAGG	CCAGGCCAACGCGTGAACATCATGATGACAGAGCA	ATCTTC	1133	
Db	1081	CTGGCCGAGCGCGATGAGG	CCAGGCCAACACCC---ACCGTGAATGATGACAGAGCA	ATCTTC	1133	
QY	1138	AAAGGCGCCCGGCGCAACGT	CAAGTGTCTTCAATCTGCGGCAAGAGAGGCGCACATCGCCAG	1197		
Db	1138	AAAGGCGCCCGGCGCATGT	CAAGTGTCTTCAATCTGCGGCAAGAGAGGCGCACATCGCCGCG	1199		
QY	1198	AACTGCGCGCGCCCCCGGCA	AGAGGCGTGTGGAAGTCCGGCAAGAGGCGCACCAATG	1255		
Db	1198	AACTGCGCGCGCCCCCGGCA	AGAGGCGTGTGGAAGTCCGGCAAGAGAGGCGCACCAATG	1255		
QY	1258	AAAGACTGCAACGAGCGC	CAAGGCCCACTTCTTGAGCAAGATCTG	GGCCCAAGCCCAAGAGGC	1317	
Db	1258	AAAGACTGCAACGAGCGC	CAAGGCCCACTTCTTGAGCAAGATCTG	GGCCCAAGCCCAAGAGGC	1317	
QY	1318	CGCCCGGCACTTCTTG	CAAGAACCGACGAGCGCGCGCC	CCACCGTGGCCACCGC	1377	



Db 1318 CCCCCCGGCACTTCTGAGAGCCG-----CCCCGAGCCACCGCC 1359  
Qy 1378 CCCCCCGGAGAGCTTCGCTTGAAGAGACCCCGCCCAAGAGAGCCCAAG 1437  
Db 1360 CCCCCCGGAGAGCTTCGCTTGAAGAGACCCCGCCCAAGAGAGAGCAAG 1419  
Qy 1438 GACCGGAGAGCCCTGACCGGAGCCCTGAGCCCTGCGAGCCCTGTTCCGAGCCGCC 1497  
Db 1420 G-----ACCGGAGAGCCCTGACCGCTGAGAGCCCTGTTCCGAGCAAGACCC 1467  
Qy 1498 CTGAGCCAGTAA 1509  
Db 1468 CTGAGCCAGTAA 1479

RESULT 10  
AX468547 1479 bp DNA 1linear PAT 16-JUL-2002  
LOCUS Sequence 67 from Patent WO0226209.  
DEFINITION AX468547  
ACCESSION AX468547  
VERSION AX468547.1 GI:21901377  
KEYWORDS  
SOURCE Human immunodeficiency virus 1 (HIV-1)  
ORGANISM Human immunodeficiency virus 1  
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate  
lentivirus group.  
REFERENCE 1  
AUTHORS O'Hagan,D., O'teen,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
Singh,M., Ulmer,J. and Dubensky,J.W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 67 04-APR-2002;  
CHIRON CORPORATION (US)  
FEATURES  
Location/Qualifiers  
1..1479 /organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"

ORIGIN  
Query Match 84.2%; Score 1271.2; DB 6; Length 1479;  
Best Local Similarity 91.8%; Pred. No. 3.6e-125;  
Matches 1388; Conservative 10; Mismatches 78; Indels 36; Gaps 4;

Qy 1 ATGGGGGCGCGGCGGAGCATCTCGCGCGGCGGAGAGCTGAGACAGTGGAGAAATCCGC 60  
Db 1 ATGGGGGCGCGGCGGAGCATCTCGCGCGGCGGAGAGCTGAGACAGCTGGAGGCGATCCGC 60  
Qy 61 CTGGCGCGCGCGGAGAGAGCATCATGCTGAAGACCTGTGTGGGCGAGCGCGAG 120  
Db 61 CTGGCGCGCGCGGAGAGAGTCTACATGATGAAGACCTGTGTGGGCGAGCGCGAG 120  
Qy 121 CTGGAGGGGCTTGGCTCTGAACCCCGGCTGTGAGAGACCGCGAGGGCTGCAAGCATC 180  
Db 121 CTGGAGAAATTGCGCTGAACCCCGGCTGTGAGAGACCGAGGGGCTGCAAGCATC 180  
Qy 181 ATGAAGAGCTGAGCGCGCGCTGAGAGACCGGAGACCGAGAGCTGGAGCTGTAAAC 240  
Db 181 ATCCCGCAGCTGACCCCGCTTGAAGACCGGAGAGAGAGCTGAAGAGCTGTAAAC 240  
Qy 241 ACCGTGGCACCCTGTACTGTGACGCGCGGATCGAGTCCCGGAGACCAAGAGGCC 300  
Db 241 ACCGTGGCACCCTGTACTGTGACGAGAGAGAGAGTGAAGTSCGAGACCAAGAGGCC 300  
Qy 301 CTGGAGAGATGAGAGAGAGAGAAAGTCCAGAGAGAGACCGAGAGGCCAAGAG 360  
Db 301 CTGGAGAGATGAGAGAGAGAGAAAGTCCAGAGAGATCAGAGAGGCCAAGAG 360  
Qy 361 GCCGAC-----GGCAGGTGAGCAGAACTACCCATCGAGAGAACTTCAGAGGCCAAG 417  
Db 361 GCCGACAAAGGAGAGGTGAGCAGAACTACCCATCGAGAGAACTTCAGAGGCCAAG 420  
Qy 418 GTGACCAAGGCGATGAGCCCGGACCTGTGAAGCGCTGAGTGAAGTATGAGAGAG 477

Db 421 GTGACCAAGGCGATGAGCCCGGACCTGTGAAGCGCTGAGTGAAGTATGAGAGAG 480  
Qy 478 GCGTTCAAGCCCGAGAGTATCCCATGTTCAACCGGCTGAGAGAGAGGCCACCCCGAG 537  
Db 481 GCTTCAAGCCCGAGAGTATCCCATGTTCAACCGGCTGAGAGAGAGGCCACCCCGAG 540  
Qy 538 GACCTGAACAGATGTTGAACAACGTTGGCGGCGCAACAGGCCCGCATGAGTGTGAAG 597  
Db 541 GACCTGAACAGATGTTGAACAACGTTGGCGGCGCAACAGGCCCGCATGAGTGTGAAG 600  
Qy 598 GACACCATCAAGAGAGAGCGCGAGTGGAGCGGCTGACACCCCGTGAAGGCCGCCGCC 657  
Db 601 GACACCATCAAGAGAGAGCGCGAGTGGAGCGGCTGACACCCCGTGAAGGCCGCCGCC 660  
Qy 658 GTGGCCCGCGCAGATCGGACCCCGCGGAGAGACATGCGCGGCGCAACAGAC 717  
Db 661 ATCGCCCGCGCAGATCGGAGCGCGCGGAGAGACATGCGCGGCGCAACAGAC 720  
Qy 718 CTGAGAGAGCAGATGCTGTGATGACAGCAACCTCCCGTGCCTGTGGGAGCATTTAC 777  
Db 721 CTGAGAGAGCAGATGCTGTGATGACAGCAACCTCCCATGCCGTGGGAGCATTTAC 780  
Qy 778 AAGCGGTGATCATCTGGGCGTGAACAAGATCGGCGGATGTACAGCCCGGTGAGCATC 837  
Db 781 AAGCGGTGATCATCTGGGCGTGAACAAGATCGGCGGATGTACAGCCCGGTGAGCATC 840  
Qy 838 CTGACATCCGCGCAGAGGCCCGCAAGAGGCTTTCGCGCATGACAGCCGCTTTCAG 897  
Db 841 CTGACATCAAGAGAGGCCCGCAAGAGGCTTTCGCGCATGACAGCCGCTTTCAG 900  
Qy 898 ACCCTGGCGCGGAGAGAGGCCCAACCAAGAGCTGAAGAACTGAGTGAACCGCATCTGTG 957  
Db 901 ACCCTGGCGCGGAGAGAGGCCCAACCAAGAGCTGAAGAACTGAGTGAACCGCATCTGTG 960  
Qy 958 GTGAGAGAGCAGACCCCGCATGCAAGACATCTCGCGCGCTTCGCGCGCGCGCGCAC 1017  
Db 961 GTGAGAGAGCAGACCCCGCATGCAAGACATCTCGCGCGCTTCGCGCGCGCGCGCAC 1020  
Qy 1018 CTGAGAGAGATGATGACCGCTGCGAGAGGCGTGGGCGCGCGCGCGCACAGAGCGCGCTG 1077  
Db 1021 CTGAGAGAGATGATGACCGCTGCGAGAGGCGTGGGCGCGCGCGCGCACAGAGCGCGCTG 1080  
Qy 1078 CTGGCGGAGGGGATGAGAGAGAGCCCAAGCCTGAACATCATGATGAGAGAGAGCAATTC 1137  
Db 1081 CTGGCGGAGGGGATGAGAGAGAGCCCAAGCCTGAACATCATGATGAGAGAGAGCAATTC 1137  
Qy 1138 AAGGGCGCGCGCGCAAGTCAAGTCTTCACTGCGGCAAGAGAGGCCACATCGCCAG 1197  
Db 1138 AAGGGCGCGCGCGCGCATGTAAGTCTTCACTGCGGCAAGAGAGGCCACATCGCCAG 1197  
Qy 1198 AACTGCGCGCGCGCGCGAGAGAGGCTGTGAAGTGGCGCAAGAGAGGCCACAGATG 1257  
Db 1198 AACTGCGCGCGCGCGCGAGAGAGGCTGTGAAGTGGCGCAAGAGAGGCCACAGATG 1257  
Qy 1258 AAGGACTGACAGAGAGCGCAGAGCCCACTTCCTGGGCAAGATCTGGCCAGCAAGAGGC 1317  
Db 1258 AAGGACTGACAGAGAGCGCAGAGCCCACTTCCTGGGCAAGATCTGGCCAGCAAGAGGC 1317  
Qy 1318 CGCCCGGCGCACTTCTGAGAGAACGAGAGCGAGCGCGCGCGCGCGCACCTGTGCCAGCGCC 1377  
Db 1318 CGCCCGGCGCACTTCTGAGAGAACGAGAGCGAGCGCGCGCGCGCGCACCTGTGCCAGCGCC 1377  
Qy 1378 CCCCCCGGAGAGCTTCGCTTGAAGAGACACCCCGCGCGCGCGCGCGAGAGAGCCAG 1437  
Db 1360 CCCCCCGGAGAGCTTCGCTTGAAGAGACACCCCGCGCGCGCGCGCGAGAGAGAG 1419  
Qy 1438 GACCGGAGAGCCCTGACCGGAGCCCTGACCGCCCTGTGGAGAGCTTTCGAGAGGCC 1497  
Db 1420 G-----ACCGGAGAGCCCTGACCGCTGTGAAGAGCTTTCGAGAGAGACCC 1467  
Qy 1498 CTGAGCCAGTAA 1509





Viruses; Retroviral viruses; Retroviridae; Lentivirus; Primate  
Lentivirus group.

REFERENCE  
1 Shao, Y., Wagner, R., Wolf, H. and Graf, M.  
The genome of the hiv-1 inter-subtype (c/b') and use thereof  
Patent: WO 0136614-A 2 25-MAY-2001;  
Genent GMBH Gesesellschaft fuer angewandte Biotechnologie (DB) ;  
Shao, Yiming (CN)

FEATURES  
source  
1. .4288  
/organism="Human Immunodeficiency Virus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:12721"

## ORIGIN

Query Match 79.7%; Score 1202.2; DB 6; Length 4288;  
Best Local Similarity 88.5%; Pred. No. 4.9e-118;  
Matches 1335; Conservative 0; Mismatches 153; Indels 21; Gaps 2;

```
QY 1 ATGGGGCGCCGCGCGCATCTCGCGCGCGGAGAGCTGAGCAAGTGGAGAGAAATCCGC 60
Db 13 ATGGGCGCGCGCGCGCATCTCGAGGGCGCGAGCTGAGCAAGTGGAGAGAAATCCAG 72
QY 61 CTGCGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGCGCAGCGCGAG 120
Db 73 CTGAGGCGCGCGCGCGCAAGAGCACTACATGCTGAAGCACTGTGTGGCGCAGCGAG 132
QY 121 CTGAGAGGCTTGGCTTGAACCCCGCGCTGTGTGAAGACCGCGAGGGCTGACAGCATC 180
Db 133 CTGAGAGAGGTTGGCTTGAACCCCGCGCTGTGTGAAGACCGAGGGCTGCAAGAGATC 192
QY 181 ATGAAGCAGCTCAGCGCGCGCGCGCGAGACCGGAGACCGAGAGAGCTGGATCAAC 240
Db 193 ATGAAGCAGCTCAGAGCGCGCGCGCGCGCGCGAGACCGAGAGAGCTGGATCAAC 252
QY 241 ACCGTGGCGCACCTTGTACTGTGTGCAACCGCGCGCATGTAGGTCCGCGACCAAGAGAGCC 300
Db 253 ACCGTGGCGCACCTTGTACTGTGTGCAACCGAGATGTAGGTGAGAGAGAGCTGTTCAC 312
QY 301 CTGGAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 313 CTGGAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 372
QY 361 GCGGACGCGAGAGTGAAGCAAGTACCCTGTGTGAAGAGAGAGAGAGAGAGAGAGAG 420
Db 373 GCGGACGCGAGAGTGAAGCAAGTACCCTGTGTGAAGAGAGAGAGAGAGAGAGAGAG 432
QY 421 CACGAGGCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 433 CACGAGGCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492
QY 481 TTACGCGCGCGAGTATCCCATTTTCAACCGCGCGCGAGAGAGAGAGAGAGAGAGAG 540
Db 493 TTACGCGCGCGAGTATCCCATTTTCAACCGCGCGCGAGAGAGAGAGAGAGAGAGAG 552
QY 541 CTGAACAAGATGTTGAACAACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 553 CTGAACAACAATGTGAACAACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
QY 601 ACCATCAACGAGAGAGCGCGCGAGTGGAGCGCGCTGCAACCGCGCGCGAGCGCGCG 660
Db 613 ACCATCAACGAGAGAGCGCGCGAGTGGAGCGCGCTGCAACCGCGCGCGAGCGCGCG 672
QY 661 GCGCGCGCGCGAGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Db 673 GCGCGCGCGCGAGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 732
QY 721 CAGAGCAGATCGCTGTGTATGACCAAGCAACCGCGCGCGCGCGCGCGCGCGCGCG 780
Db 733 CAGAGCAGATCGCTGTGTATGACCAAGCAACCGCGCGCGCGCGCGCGCGCGCGCG 792
QY 781 CGGTGATCATCTTGGCGCTGAAACAAGTGTGGAGATGACAGCGCGCGCGCGCGCG 840
```

```
Db 793 AGTGGATCATCTTGGGTTTAAACAAGATCGTGAAGATGTACAGCCCAACCAAGATCCTG 852
QY 841 GACATCCGCGCAGAGGCGCGCGAGAGAGCGCTTCCGCGCATACGTGAGACCGCTTCTTCAAGAC 900
Db 853 GACATCAAGCAGAGGCGCGCGAGAGAGCGCTTCAAGGAGCTACGTGACAGAGTTCTTCAAGAC 912
QY 901 CTGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Db 913 CTGAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 972
QY 961 CAGAAAGCGCAACCGCGCATGCAAGAGCACTCTGTGCGCGCTCGGCGCGCGCGCGCGCG 1020
Db 973 CAGAAAGCGCAACCGCGCATGCAAGAGCACTCTGTGAGGCGCTGGGCGCGCGCGCGCGCG 1032
QY 1021 GAGAGATGATGACCGCTGTGCGAGGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Db 1033 GAGAGATGATGACCGCTGTGCGAGGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1092
QY 1081 GCGGAGCGAGTGAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1093 GCGGAGCGCATGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1149
QY 1141 GCGCGCGCGCGCAACGTCAAGTGTCTTCACTGCGCGCAAGAGAGAGAGAGAGAGAGAG 1200
Db 1150 GCGAGCAAGAGAGATGTGAAGTGTCTTCACTGCGCGCAAGAGAGAGAGAGAGAGAGAG 1209
QY 1201 TCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1210 TCGAGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1269
QY 1261 GACTGCAACGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1270 GACTGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1329
QY 1321 CCGCGCAACTTCTGCAAGAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1330 CCGCGCAACTTCTGCAAGAACG-----GCCGAGGCCACCGCGCCC 1371
QY 1381 CCGCGCGAGAGCTTCCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1372 CCGGAGAGAGCTTCAAGTTCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
QY 1441 GCGGAGCGCTTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1432 ATCAACAAGAGAGCTGTACCTCTGACAGCTGTGAAGAGCTGTTCGCGCAAGACCTCAAGC 1491
QY 1501 AGCCAGTAA 1509
Db 1492 AGCCAGTAA 1500

RESULT 13
AF201927 1509 bp DNA linear SYN 16-MAR-2000
LOCUS AF201927
DEFINITION Synthetic construct gag protein gene, complete cds.
ACCESSION AF201927
VERSION AF201927.1 GI:7248702
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 1509)
AUTHORS zur Megede, J., Chen, M. C., Doe, B., Schaefer, M., Greer, C. E.,
Selby, M., Otten, G. R. and Barnett, S. W.
TITLE Increased expression and immunogenicity of sequence-modified human
JOURNAL J. Virol. 74 (6), 2628-2635 (2000)
MEDLINE 20148954
PUBMED 10684277
REFERENCE
2 (bases 1 to 1509)
AUTHORS zur Megede, J. and Barnett, S. W.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1999) Vaccines, Chiron Corporation, 4560 Horton,
```



Qy	1	ATGGGCGCCGCGCCAGCATCTCTGGCGCGGCGAGAACTTGGAACAATGGGAGAAATGATCCGC	60
Db	7	ATGGGCGCGCCCGCGCAGCGTGTGAGCGGCGCGAGCTTGGAACAATGGGAGAAATGATCCGC	66
Qy	61	CTGGGCGCCCGCGCGGCAAGAACCTAACATGCTGGAAGCACTGTGATGTGGCCCGCGCGAG	120
Db	67	CTGGGCGCCCGCGCGGCAAGAAATGACAAAGCTGAAGCAATGTGTGGCCCGCGCGAG	126
Qy	121	CTGAGAGGGCTTCGCCCCTGAACCCCGGCGCTGTGAGAACCGCGGAGGCTGCAAGCAGATC	180
Db	127	CTGAGAGCGCTTCGCCCCTGAACCCCGGCGCTGTGAGAGCAAGGAGGGCTGCGCCAGATC	186
Qy	181	ATGAAGCAGCTGCGAGCCCGCCCTTGAGAACCGGCAACGGAGAGCTGTGCGAGCTGTAAACA	240
Db	187	CTGGGCGCAGCTGCGAGCCCGCCCTTGAGAACCGGCAAGAGAGACTGTGCGAGCTGTAAACA	246
Qy	241	ACCGTGCACACTCTGTACTGCGTGCAGCGCCGGCACTGAGGTCGCGCAACAAGAGAGCC	300
Db	247	ACCGTGGCGCACTCTGTACTGCGTGCAGCGCCATCGACTGAAGAACACCAAGAGAGCC	306
Qy	301	CTGACAAGATTCAGAGAGAGAGCAGAAACAATGCCCAAGAAAGACCACAGAGGCCAAGAG	360
Db	307	CTGAGAAGATTCAGAGAGAGAGCAGAAACAAGTCCAAAGAAAGAGGCCCAAGAGGCCGCC	366
Qy	361	GCCGACGGCA-----AGGTGAGCCAGAACTAACCCCATCTGTGAGAACCTG	405
Db	367	GCTCCGCGCAACCGGCAACAGCAGCCAGGTGAGCCAGAACTAACCCCATGTGTGAGAACCTG	426
Qy	406	CAGGCGCAGATGTGTGACACAGGCCATCAGGCCCGGCAACCTTAAGACGCTGGGGTGAAGGTG	465
Db	427	CAGGCGCAGATGTGTGACACAGGCCATCAGGCCCGGCAACCTTAAGACGCTGGGGTGAAGGTG	486
Qy	466	ATCAGAGAGAAAGGCTTCAGCCCGGCGAGGTATTCCTCATGTTCAACGCCCTTGACGAGGAC	525
Db	487	GTGAGAGAGAAAGGCTTCAGCCCGGCGAGGTATTCCTCATGTTCAAGGCCCTTGACGAGGAC	546
Qy	526	GCCACCCCCCGCAGACTGAACAACATGTTGAACACCTGGGCGGCGCACAGAGCGCCATG	585
Db	547	GCCACCCCCCGCAGACTGAACAACATGTTGAACACCTGGGCGGCGCACAGAGCGCCATG	606
Qy	586	CAGATGCTGAAGGACACCATCAACAGAGAGGCCGCGGAGTGGAGACCGGCTTGACCCCGTG	645

Db	607	CAGATGCTGAAGAGACCATCAACAGAGAGGCGCCGGAATGGGAACCGCGTGACCCCGTG	668
Qy	646	CAGCGCGGCCCCCGTGGCCCCCGGCGCAGATGCGGACCCCCCGGCGCAGCGACATCGCCGGC	705
Db	667	CAGCGCGGCCCCCATCGCCCCCGGCGAGATGCGCGAGCCCCCGGCGCAGCGACATCGCCGGC	726
Qy	706	GCACACAGCAACCTGTGAGAGAGAGATCGCTGGAATGACCAAGCAACCCCCCGGCGCGTG	765
Db	727	AACCAACAGAACCTTCGAGAGAGAGATCGGCTGGATGACCAACACCCCCCATCCCCGTG	786
Qy	766	GGCGCATTTACAGCGGTGATCATCTTGCGCTTGAAACAAGTCGTGCGATGTACACG	825
Db	787	GGCGGATTTACAGCGGTGATCATCTTGCGCTTGAAACAAGTCGTGCGATGTACACG	846
Qy	826	CCCGGAGCATCTTGGACATCCGCGAAGGCCCCAAGAGGCCCTTCGCGGACATACGTGAGAC	885
Db	847	CCCAACAGCATCTTGGACATCCGCGAAGGCCCCAAGAGGCCCTTCGCGGACATACGTGAGAC	906
Qy	886	CGCTTCTTCAAGACCCCTGCGCGCCGAGCAGGCAACCCCAAGACGTGAAGATCTGATGACC	945
Db	907	CGCTTCTTCAAGACCCCTGCGCGCTGAGCAGGCAACCCCAAGACGTGAAGATCTGATGACC	966
Qy	946	GAGACCTCTGTGTGTGACAAAGGCCCAACCCCACTGTGAAGACATCTCTGCGCTCTCGGC	1001
Db	967	GAGACCTCTGTGTGTGACAAAGGCCCAACCCCACTGTGAAGACATCTCTGCGCTCTCGGC	1022
Qy	1006	CCCGGCGCCACCTCTGAGAGAGATGATGACCGGCTCTGACAGGCGGTGAGCGGCCCGGCGCAC	1061
Db	1027	CCCGGCGCCACCTCTGAGAGAGATGATGACCGGCTCTGACAGGCGGTGAGCGGCCCGGCGCAC	1081
Qy	1066	AAGGCGCGCGTCTGTGCGCGAGGCGCATGAGCCAGG---CCACAGCGTGAACATCATGATG	1122
Db	1087	AAGGCGCGCGTCTGTGCGCGAGGCGCATGAGCCAGG---CCACAGCGTGAACATCATGATG	1144
Qy	1123	CAGAAAGCAACTTGAAGGCGCCCCCGGCGCAACGTCAAGTGTCTTGAACCTGCGCGCAAGAG	1184
Db	1147	CAGCGCGGCACTTCCGCAACCAAGCGGAAGACCGTCAAGTGTCTTGAACCTGCGCGCAAGAG	1201
Qy	1183	GGCCACATCGCCAGAACCTGCGCGCGCCCCCGCAAGAAAGGCGTGTGAAGTGCAGAG	1244
Db	1207	GGCCACATCGCCAGAACCTGCGCGCGCCCCCGCAAGAAAGGCGTGTGAAGTGCAGAG	1266
Qy	1243	GAGGGCCACAGATGAAGGACCTGACCGAGAGGCGCAAGGCACTTCTGCGGCAAGATCTGCG	1304
Db	1267	GAGGGCCACAGATGAAGGACCTGACCGAGAGGCGCAAGGCACTTCTGCGGCAAGATCTGCG	1322
Qy	1303	CCCAAGCCAAAGGCGCGCGCCCGGCAACTTCTCTGCGAAGACCGAGCGAAGCCCGCCGCCCC	1367
Db	1327	CCCAAGCCAAAGGCGCGCGCCCGGCAACTTCTCTGCGAAGACCGAGCGAAGCCCGCCGCCCC	1368
Qy	1363	AACGTGCCCCACCGCCCCCGCGCGAGAGCTTTCGCTTTCGAGAGAACCAACCCCGCGCCCC	1422
Db	1389	CCCAAGCCCAACCGCCCCCGCGAGAGCTTTCGCTTTCGAGAGAGAAAGCAACCCCGC	1422
Qy	1423	AAGCAGAGCCCAAGAGACCGGAGCGCTTACCGCGAGACCCCTGACCGCGCTGCGCAGCTG	1487
Db	1429	AGCCAGAGAGAGAGGCCATCGAACAGGAGCGTGAACCCCTGACCAAGCTGCGCAGCTG	1488
Qy	1483	TTTCGCGACGCGCCCTTGAGCCAGTTAA	1509
Db	1489	TTTCGCGACGACCCCAAGCAGCCAGTTAA	1515

RESULT 15			
CO870496			
LOCUS	CO870496	1515 bp	DNA
DEFINITION	Sequence 4 from Patent EPI433851.		linear
ACCESSION	CO870496		
VERSION	CO870496.1	GI:5200023	
KEYWORDS	.		
SOURCE	synthetic construct		
ORGANISM	synthetic construct		
	other sequences; artificial sequences.		

REFERENCE 1  
AUTHORS Barnett, S., Zurmege, J., Srivastava, I., Lian, Y., Hartog, K.,  
TITLE Liu, H., Greer, C., Selby, M., and Walker, C.  
JOURNAL Improved expression of HIV polyprotein and production of  
virus-like particles  
Patent: EP 143851-A 4 30-JUN-2004;  
CHIRON CORPORATION (US)  
FEATURES  
source Location/Qualifiers  
1.1515  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Description of Artificial Sequence: synthetic  
HIV-gag"  
ORIGIN  
Query Match 79.6%; Score 1201; DB 6; Length 1515;  
Best Local Similarity 88.5%; Pred. No. 8.8e-118;  
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;  
QY 1 ATGGGGCCCCCGCCAGCATCTTCGGCGGAGAGAGTGAAGTGGAGAAATCCGC 60  
DB 7 ATGGGGCCCCCGCCAGCATCTTCGGCGGAGAGAGTGAAGTGGAGAAATCCGC 66  
QY 61 CTGGCGCCCGCGGCAAGAAAGCACTAATGCTGAAGCACTGGTGTGGGCGAGCGGAG 120  
DB 67 CTGGCGCCCGCGGCAAGAAAGTAACTAAGCTAAGCACTGTTGGGCGAGCGGAG 126  
QY 121 CTGGAGGGGCTTCGCTCTGAGACCCCGGCTGCTGAGAGCCGCGAGGGCTGCAAGATC 180  
DB 127 CTGGAGCGCTTGGCGGTGAACCCCGGCTGCTGAGAGCCGAGGGGCTGCGCGAGATC 186  
QY 181 ATGAGACAGCTGAGACCCCGCTGAGACCGGCACTGGAGAGCTGGGCGAGCTGTACAC 240  
DB 187 CTGGGCGACCTGAGACCCCGCTGAGACCGGCACTGGAGAGCTGGGCGAGCTGTACAC 246  
QY 241 ACCGTGGCCACCTGTACTGCTGAGACCGCGGATGAGGTCCGCGACCAAGAGAGGCC 300  
DB 247 ACCGTGGCCACCTGTACTGCTGAGACCGCGGATGAGGTCAAGAGACCAAGAGAGGCC 306  
QY 301 CTGGAAGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGGAAGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCGGAGAGCA-----AGTGAAGCAAGAACTACCCCTCTGTGAGAACTTG 405  
DB 367 GCGGCGGAGACCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426  
QY 406 CAGGGCCAGATGAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465  
DB 427 CAGGGCCAGATGAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486  
QY 466 ATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525  
DB 487 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546  
QY 526 GCCACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585  
DB 547 GCCACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606  
QY 586 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 645  
DB 607 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666  
QY 646 CAGGCGGCGCGCTGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 705  
DB 667 CAGGCGGCGCGCGCTGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
QY 706 GCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765  
DB 727 ACCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786  
QY 766 GCGGACATGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825

DB 787 GCGGAGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 846  
QY 826 CCGGTAGAGATCTGAGAGATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885  
DB 847 CCACAGAGATCTGAGAGATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906  
QY 886 CGCTTCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 945  
DB 907 CGCTTCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966  
QY 946 GAGACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005  
DB 967 GAGACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026  
QY 1006 CCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1065  
DB 1027 CCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086  
QY 1066 AAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1122  
DB 1087 AAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146  
QY 1123 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1182  
DB 1147 CAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206  
QY 1183 GCGGACATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1242  
DB 1207 GCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266  
QY 1243 GAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302  
DB 1267 GAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326  
QY 1303 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1362  
DB 1327 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368  
QY 1363 ACCGTGGCCACCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1422  
DB 1369 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1428  
QY 1423 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482  
DB 1429 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488  
QY 1483 TTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1509  
DB 1489 TTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1515

Search completed: March 12, 2005, 21:18:23  
Job time : 6741.44 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 16:41:47 ; Search time 266.651 Seconds

(without alignments)  
9259.839 Million cell updates/sec

Title: US-09-475-704a-4

Perfect score: 1509

Sequence: 1 atggggccggccggccagcatc.....ggggcccttgagccagctaa 1509

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405566

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/pdata/1/ina/5A.COMB.seq:\*

2: /cgn2\_6/pdata/1/ina/5B.COMB.seq:\*

3: /cgn2\_6/pdata/1/ina/6A.COMB.seq:\*

4: /cgn2\_6/pdata/1/ina/6B.COMB.seq:\*

5: /cgn2\_6/pdata/1/ina/6C.COMB.seq:\*

6: /cgn2\_6/pdata/1/ina/6D.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1201	79.6	1515	4	US-09-475-515-4
2	1201	79.6	4472	4	US-09-475-515-75
3	1201	79.6	4608	4	US-09-475-515-76
4	1201	79.6	4689	4	US-09-475-515-74
5	1201	79.6	4766	4	US-09-475-515-73
6	1199	79.5	2031	4	US-09-475-515-7
7	1117.8	74.1	1853	4	US-09-475-515-5
8	1117.8	74.1	1865	4	US-09-475-515-78
9	1117.8	74.1	1865	4	US-09-475-515-79
10	1117.8	74.1	4319	4	US-09-475-515-6
11	1049.2	69.5	1268	4	US-09-475-515-9
12	971.4	64.4	1521	4	US-09-952-060-27
13	971.4	64.4	1532	4	US-09-974-702-1
14	971.4	64.4	1532	4	US-09-818-443-1
15	971.4	64.4	37474	4	US-09-952-060-25
16	968.4	64.2	4053	4	US-09-952-060-34
17	930.2	61.6	4307	4	US-09-936-572-2
18	929.4	61.6	4307	4	US-09-952-950-2
19	925.4	61.3	4327	4	US-09-936-572-14
20	925.4	61.3	4353	4	US-09-936-572-13
21	925.4	61.3	4642	4	US-09-936-572-12
22	925.4	61.3	9772	4	US-09-552-950-5
23	916.6	60.7	1482	4	US-09-818-443-4
24	913.6	60.5	1479	4	US-09-952-060-32
25	736.6	48.8	1496	4	US-09-872-733A-6
26	734.2	48.7	1496	4	US-09-184-418C-82
27	734.2	48.7	8972	4	US-09-184-418C-9

28	730.4	48.4	1476	4	US-09-991-258-4	Sequence 4, Appl
29	730.4	48.4	12523	4	US-09-991-258-1	Sequence 1, Appl
30	723.8	48.0	4338	4	US-09-872-733A-1	Sequence 1, Appl
31	712.8	47.2	9913	4	US-09-827-688-11	Sequence 11, Appl
32	709.6	47.0	1485	4	US-09-184-418C-73	Sequence 73, Appl
33	709.6	47.0	9010	4	US-09-184-418C-8	Sequence 8, Appl
34	708	46.9	1476	4	US-09-184-418C-100	Sequence 100, Appl
35	708	46.9	8959	4	US-09-184-418C-11	Sequence 11, Appl
36	663.6	44.0	1486	4	US-09-184-418C-38	Sequence 38, Appl
37	663.6	44.0	8992	4	US-09-184-418C-4	Sequence 4, Appl
38	651.8	43.2	8968	4	US-09-184-418C-12	Sequence 12, Appl
39	649.4	43.0	1479	4	US-09-184-418C-1	Sequence 1, Appl
40	642.2	42.6	7399	2	US-08-418-848A-9	Sequence 9, Appl
41	642.2	42.6	9709	2	US-08-188-583-5	Sequence 5, Appl
42	642.2	42.6	9709	3	US-08-388-353-1	Sequence 1, Appl
43	642.2	42.6	9709	3	US-08-488-551B-1	Sequence 1, Appl
44	642.2	42.6	9709	3	US-09-309-572-15	Sequence 15, Appl
45	642.2	42.6	9709	4	US-09-718-096-15	Sequence 15, Appl

## ALIGNMENTS

RESULT 1  
US-09-475-515-4  
Sequence 4, Application US/09475515A

Patent No. 6602705

GENERAL INFORMATION:

APPLICANT: BARNETT, Susan

APPLICANT: ZUR MEGEDE, Jan

APPLICANT: SRIVASTAVA, Indresh

APPLICANT: LIAN, Ying

APPLICANT: HARTOG, Karin

APPLICANT: LIU, Hong

APPLICANT: GREER, Catherine

APPLICANT: SELBY, Mark

APPLICANT: WALKER, Christopher

TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION

TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES

FILE REFERENCE: 1621.002

CURRENT APPLICATION NUMBER: US/09/475,515A

CURRENT FILING DATE: 1999-12-30

NUMBER OF SEQ ID NOS: 90

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 4

LENGTH: 1515

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

US-09-475-515-4

Query Match 79.6%; Score 1201; DB 4; Length 1515;  
Best Local Similarity 88.5%; Pred No. 2.8e-173; Indels 36; Gaps 3;  
Matches 1351; Conservative 0; Mismatches 140;

QY	1	ATGGGCGCCCGCGCGCATCTGGCGCGGAGAGCAAGCTGACCAAGTGGAGAAATCCGC	60
DB	7	ATGGGCGCCCGCGCGCATCTGGCGCGGAGAGCAAGTGGAGAAATCCGC	66
QY	61	CTGGGCGCCCGCGCGGAGAGCAAGTGGAGAAATCCGC	120
DB	67	CTGGGCGCCCGCGCGGAGAGCAAGTGGAGAAATCCGC	126
QY	121	CTGGGCGCCCGCGCGGAGAGCAAGTGGAGAAATCCGC	180
DB	127	CTGGGCGCCCGCGCGGAGAGCAAGTGGAGAAATCCGC	186
QY	181	ATGAAGCAGTGCAGCCCGCGCGGAGAGCAAGTGGAGAAATCCGC	240
DB	187	CTGGGCGCGGAGAGCAAGTGGAGAAATCCGC	246

241 ACCGTGGCCACCTGTACTGTCAGCGCGGATCGAGGTCCGCGACCAAGAGAGCC 300  
247 ACCGTGGCCACCTGTACTGTCAGCGCGGATCGAGGTCCGCGACCAAGAGAGCC 306  
301 CTGGAACAAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
307 CTGGAACAAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
361 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405  
367 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426  
406 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465  
427 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486  
466 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525  
487 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546  
526 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585  
547 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606  
586 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 645  
607 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666  
646 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 705  
667 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
706 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765  
727 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786  
766 GCGAGATCTTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825  
787 GCGAGATCTTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 846  
826 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885  
847 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906  
886 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 945  
907 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966  
946 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005  
967 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026  
1006 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1065  
1027 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086  
1066 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1122  
1087 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146  
1123 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1182  
1147 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206  
1183 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1242  
1207 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266  
1243 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302  
1267 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326  
1303 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1362

1327 CCAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368  
1363 ACCGTGGCCACCTGTACTGTCAGCGCGGATCGAGGTCCGCGACCAAGAGAGCC 1422  
1369 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1428  
1423 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482  
1429 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488  
1483 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1509  
1489 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1515

RESULT 2  
US-09-475-515-75  
Sequence 75, Application US/09475515A  
GENERAL INFORMATION:  
PATENT NO. 6602705  
APPLICANT: BARNETT, Susan  
APPLICANT: ZUR MEGHEDE, Jan  
APPLICANT: SRIVASTAVA, Indresh  
APPLICANT: LIAN, Ying  
APPLICANT: HARTOG, Karin  
APPLICANT: LIU, Hong  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
FILE REFERENCE: 1621.002  
CURRENT FILING DATE: 1999-12-30  
CURRENT APPLICATION NUMBER: US/09/475, 515A  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 75  
LENGTH: 4472  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
US-09-475-515-75

Query Match 79.6%; Score 1201; DB 4; Length 4472;  
Best Local Similarity 80.5%; Pred. No. 2.8e-173;  
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;  
1 ATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
2339 ATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2398  
61 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
2299 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3058  
121 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
3059 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3118  
181 ATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
3119 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3178  
241 ACCGTGGCCACCTGTACTGTCAGCGCGGATCGAGGTCCGCGACCAAGAGAGCC 300  
3179 ACCGTGGCCACCTGTACTGTCAGCGCGGATCGAGGTCCGCGACCAAGAGAGCC 3238  
301 CTGGAACAAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
3239 CTGGAACAAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3298



OY	351	GGCGAGGCA-----AGGTGACCGAACTACCCCATGTTGAGAACCTG	405
Db	3259	GGCGGGGACCGGGACAGACGACCGAGGTGAGCCAGAACTACCTATGTTGAAACTTG	3358
OY	406	CAGGGCCAGATGTTGCACCAAGGCATCAGCCCCCGCACCTTGAACTGCTGGTGAAGTG	465
Db	3359	CAGGGCCAGATGTTGCACCAAGGCATCAGCCCCCGCACCTTGAACTGCTGGTGAAGTG	3418
OY	466	ATCGAGGAAAGGCTTCAGCCCCCGAGGTATCCCATGTTCAACGGCTCTAGCGAGGC	525
Db	3419	GTGAGGAGAAAGGCTTCAGCCCCCGAGGTATCCCATGTTCAAGGCTCTAGCGAGGC	3478
OY	526	GCCACCCCCCAGAGCTTGAAACAGATGTTGAAACCGTGGCGGACACAGGCGGCATG	585
Db	3479	GCCACCCCCCAGAGCTTGAAACAGATGTTGAAACCGTGGCGGACACAGGCGGCATG	3538
OY	586	CAGATGCTGAAGACACATCAAGAGAGCGCCGATGGGACCGCTTGACCCCGTG	645
Db	3539	CAGATGCTGAAGAGACATCAAGAGAGCGCCGATGGGACCGCTTGACCCCGTG	3598
OY	646	CAGGCGGGCCCCGTGGCCCCCGGGCCAGATGGCGCACCCCGCGGACAGGACATGCGCGGC	705
Db	3599	CAGGCGGGCCCCATGCCCCCGGGCCAGATGGCGCACCCCGCGGACAGGACATGCGCGGC	3658
OY	706	GCACACAGACCTCTGCAGAGGAGATGCGCTTGATGACCAAGCAACCCCGTGGCTGTG	765
Db	3659	ACCAACACACCTCTGCAGAGGAGATGCGCTTGATGACCAACCAACCCCGATCCCGGTG	3718
OY	766	GGCCACATCTTCAAGCGGTGATCTACTTGGGCTTGAAACAAGTGTGGATGTACAC	825
Db	3719	GGCCAGATCTTCAAGCGGTGATCTACTTGGGCTTGAAACAAGTGTGGATGTACAC	3778
OY	826	CCCGTGACATCTTGACATCGCGCAGAGGCCCCCAGAGACCTTTCGCGACTACGTGAC	885
Db	3779	CCCAACAGACATCTTGACATCGCGCAGAGGCCCCCAGAGACCTTTCGCGACTACGTGAC	3838
OY	886	CGCTTCTTCAAGACCTTCGCGCGCGAGCAAGGCACCCAGACGTGAGAACTGATGACC	945
Db	3839	CGCTTCTTCAAGACCTTCGCGCGCGAGCAAGGCACCCAGACGTGAGAACTGATGACC	3898
OY	946	GAGACCCCTGCGGTGCAAGCGCCACCCCGACTGCAAGACCATCTGCGGCTCTGGC	1005
Db	3899	GAGACCCCTGCGGTGCAAGCGCCACCCCGACTGCAAGACCATCTGAGAGCTTCTGGC	3958
OY	1006	CCCGGCGCACCTTGAGAGATGTGACCGCTTGCCAGGCGTGGCGCGCCCGGCAC	1065
Db	3959	CCCGGCGCACCTTGAGAGATGTGACCGCTTGCCAGGCGTGGCGCGCCCGGCAC	4018
OY	1066	AAAGCCCGGTGCTGCGCGAGGCGATGAGCCAGG---CCAAACGCTGAACATCATGATG	1122
Db	4019	AAAGCCCGGTGCTGCGCGAGGCGATGAGCCAGGCGGAGAAACCCGACACATCATGATG	4078
OY	1123	CAGAAAGCAACTTCAAGGCGCCCCCGGCGCAACGTCAAGTGTCTTCACTGCGGCAAGAG	1182
Db	4079	CAGCGCGGCAACTTCCGCAACCAAGGAGAACGCTCAAGTGTCTTCACTGCGGCAAGAG	4138
OY	1183	GGCCACATCGCCAGAACTGCGCGCGCCCCCGCAGAAAGGCGCTGTGAAGTGCAGAG	1242
Db	4139	GGCCACATCGCGCAGAACTGCGCGCGCCCCCGCAGAAAGGCGCTGTGCGCTGCGCGC	4198
OY	1243	GAGGGCCACAGATGAAGATGCACTGCAACGAGGCGCGAGCACTTCTGTTGGCAAGTCTG	1302
Db	4199	GAGGGCCACAGATGAAGATGCACTGCAACGAGGCGCGAGCACTTCTGTTGGCAAGTCTG	4258
OY	1303	CCCAAGCCCAAGGCGCGCGCGCAACTTCTGTGAAACCGCAGCGAGCCCGCGCCGCC	1362
Db	4259	CCCAAGCTTCAAGGCGCGCGCGCGCAACTTCTGTGAGACCG-----C	4300
OY	1363	ACCGTGGCAACGCGCGCGCGCGCGAGACTTTCGCTTCAAGAGACCAACCCCGCGCCC	1422
Db	4301	CCCAAGCCCAACGCGCGCGCGCGCGAGACTTTCGCTTTCGCGAGAGAGACCAACCCCG	4360

Oy	1423	AAGAGAAGCCCAAGAACCGGGAACCTTAAACGGCAGGCCCTTGACCGGCCTTGCGAGCCTG	14480
Dd	4461	AGCGAAGAGCAGAGGCCCATATGACAAGAGGTGTATCCCCCTGATCACACTCTGCGAGCCTG	4420
Oy	1483	TTCGGCAGCGGCCCCCTGAGCCACTAA	1509
Dd	4421	TTCGGAACGACGCCGAGCCACTAA	4447

RESULT 3  
US-09-475-515-76

```

; Sequence 76, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDS, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIT, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 76
; LENGTH: 4608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp160.mod$F162.delV2.gag.mod$F2
US-09-475-515-76

```

Oy	1	ATGGGCGCCCGCGCAGCATCTTGGCGGCGAGAGCTGAGCAAATGGGAGAAATCCGC	60
Dd	3075	ATGGGCGCCCGCGCAGCATGTGTGAGCGGCGGCGAGCTGGAACAAATGGGAGAAATCCGC	3134
Oy	61	CTGGCGCCCGCGCGCAAGAACATAATCTGAAGCACTGTGTGTGTGGCCAGCCCGGAG	120
Dd	3135	CTGGCGCCCGCGCGCAAGAAAGTAATMACTGAAGCAATGTGTGTGGCGAGCCCGGAG	3194
Oy	121	CTGGAGGGCTTTCCGCGCTGGAACCCCAGGCGCTGTGGAGACCGCGAGGGCTGCAAGCATC	180
Dd	3195	CTGGAGGGCTTTCCGCGGTGAACCCCAGGCGCTGTGGAGACCGCGAGGGCTGCGCAATC	3254
Oy	181	ATGAMGAGCTGACAGCCCGCGCTTGACAGCCGACCGAGAGAGCTGCGCAGCTGTACAC	240
Dd	3255	CTGGGCGCAGCTGACAGCCCGCGCTTGACAGCCGACCGAGAGAGCTGCGCAGCTGTACAC	3314
Oy	241	ACCGTGGCCACCTGTATCTGCGTGTACAGCGCGCATCGAGTCCGCAACCAAGAGAGCC	300
Dd	3315	ACCGTGGCCACCTGTATCTGCGTGTACAGCGCGCATCGAGTCCGCAACCAAGAGAGCC	3374
Oy	301	CTGGACAAAGATCGAGAGAGAGCAAGAACAGTCCAGAGAAAGACCAGAGCCAGCAAGAG	360
Dd	3375	CTGGACAAAGATCGAGAGAGAGCAAGAACAGTCCAGAGAAAGACCAGAGCCAGCAAGAG	3434
Oy	361	GCCGACGGCA-----AGGTAGCCAGAACTTACCCCATCTGTGCAAACTTG	405
Dd	3435	GCCGACGGCA-----AGGTAGCCAGAACTTACCCCATCTGTGCAAACTTG	3494
Oy	406	CAGGCGCAATGTGTGACCAAGGCATTAAGCCCCCGCACCTTGAACGCTTGGGTGAAGTG	465

Db 3495 CAGGGCCAGATGATGACACAGCCATCATGACCCCGACCTGTAAACGCTGGTGAAGTGT 3554  
Qy 466 ATCGAGAGAAAGCTTTCAGCCCCGAGGTATTCCTCCATGTTCAACCGCCCTGAGCGAGG 525  
Db 3555 GTGAGAGAAAGGCTTTCAGCCCCGAGGTATTCCTCCATGTTCAACCGCCCTGAGCGAGG 3614  
Qy 526 GCCACCCCCCAGAGCTGAAACAGATGTGAACAGCTGGGCGGCAACCGAGGCGGCAAG 585  
Db 3615 GCCACCCCCCAGAGCTGAAACAGATGTGAACAGCTGGGCGGCAACCGAGGCGGCAAG 3674  
Qy 586 CAGATGCTGAAGAACACCATCAACAGAGAGCGCGCAGTGGAGCGCTGCAACCCCGT 645  
Db 3675 CAGATGCTGAAGAACCATCAACAGAGAGCGCGCAGTGGAGCGCTGCAACCCCGT 3734  
Qy 646 CAGGCGGCGCGCTGGGCGCGCGCGAGATCGGACCCCGCGGCAAGCATGCGCGG 705  
Db 3735 CAGGCGGCGCGCTGGGCGCGCGCGAGATCGGAGCCCGCGGCAAGCATGCGCGG 3794  
Qy 706 GCCACGACGACCTGACAGAGAGATGACCTGTGATGACGAAACCCCGCGCGCGG 765  
Db 3795 ACCACGACGACCTGACAGAGAGATGACCTGTGATGACGAAACCCCGCATGCGCGG 3854  
Qy 766 GCGCATCTTACAGCGGTGATCATCTGGGCTGAAACAGATCGTGGAGTGAACG 825  
Db 3855 GCGCATCTTACAGCGGTGATCATCTGGGCTGAAACAGATCGTGGAGTGAACG 3914  
Qy 826 CCGGTGAGATCTGTGACATCCGCGCAGGCGCGCAAGAGCCCTTCCGCACTACGTGAC 885  
Db 3915 CCGGTGAGATCTGTGACATCCGCGCAGGCGCGCAAGAGCCCTTCCGCACTACGTGAC 3974  
Qy 886 CCGGTGAGATCTGTGACATCCGCGCAGGCGCGCAAGAGCCCTTCCGCACTACGTGAC 945  
Db 3975 CCGGTGAGATCTGTGACATCCGCGCAGGCGCGCAAGAGCCCTTCCGCACTACGTGAC 4034  
Qy 946 GAGACCTGTGTGAGAAACCGCAACCCGACTGCAAGACCAATCTGCGCTTCTGCG 1005  
Db 4035 GAGACCTGTGTGAGAAACCGCAACCCGACTGCAAGACCAATCTGCGCTTCTGCG 4094  
Qy 1006 CCGGCGGCGACCTGTGAGAGATGATGACGCGCTGCAAGAGGCGGCGCGCGCAC 1065  
Db 4095 CCGGCGGCGACCTGTGAGAGATGATGACGCGCTGCAAGAGGCGGCGCGCGCAC 4154  
Qy 1066 AAGGCGCGGTGTGCGGCGGAGCGATGAGCCAGG---CCAGACGCGTGAACATCATGATG 1122  
Db 4155 AAGGCGCGGTGTGCGGCGGAGCGATGAGCCAGGTGACGAAACCGGCGACCATCATGATG 4214  
Qy 1123 CAGAAAGAGACTTCAAGGCGCGCGCGGCGCAACGTCAAGTGTCTTCACTGCGCGCAAGAG 1182  
Db 4215 CAGCGCGGCACTTCCGCAACAGCGGAAAGACGTTCAAGTGTCTTCACTGCGCGCAAGAG 4274  
Qy 1183 GGCACATGCGCAAGAACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGGCGGCAAG 1242  
Db 4275 GGCACATGCGCAAGAACTGCGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGGCGGCGG 4334  
Qy 1243 GAGGCGCACAGATGAAGAGCTGCAACCGAGCGCGCAACTTCTGCGCGCAAGATCTG 1302  
Db 4335 GAGGCGCACAGATGAAGAGCTGCAACCGAGCGCGCAACTTCTGCGCGCAAGATCTG 4394  
Qy 1303 CCCAGCGCAAGAGGCGCGCGCGGCAACTTCTGCGCAAGACCGGAGGAGCCCGCGCGCC 1362  
Db 4395 CCCAGCGCAAGAGGCGCGCGCGGCAACTTCTGCGCAAGACCGGAGGAGCCCGCGCGCC 4436  
Qy 1363 ACCGTGCGCACCGCGCGCGCGCGCGAGAGCTTCCGCTTGAAGAGACCGCGCGCGCC 1422  
Db 4437 CCGGAGCGCACCGCGCGCGCGCGCGAGAGCTTCCGCTTCCGCGAGAGAGAGACCGCGCC 4496  
Qy 1423 AAGCGAGAGCGCAAGAGACCGGAGCGCTTACCGCGAGCGCGCTGAGCGCGCTGCGAGCTG 1482  
Db 4497 AAGCGAGAGCGCAAGAGACCGGAGCGCTTACCGCGAGCGCGCTGAGCGCGCTGCGAGCTG 4556  
Qy 1483 TTCCGCGAGCGCGCGCGCTGAGCGAGTAA 1509  
Db 4557 TTCCGCGAGCGAGCGCGCGAGCGCAAGTAA 4583

RESULT 4  
US-09-475-515-74  
Sequence 74, Application US/09475515A  
Patent No. 6602705  
GENERAL INFORMATION:  
APPLICANT: BARNETT, Susan  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: SRIVASTAVA, Indresh  
APPLICANT: LIAN, Ying  
APPLICANT: HARTOG, Karin  
APPLICANT: LIU, Hong  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475.515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 74  
LENGTH: 4689  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
US-09-475-515-74

Query Match 79.6%; Score 1201; DB 4; Length 4689;  
Best Local Similarity 88.5%; Pred. No. 2.8e-173;  
Matches 1351; Conservative 0; Mismatches 140; Indels 36; Gaps 3;

Qy 1 ATGGGCGCGCGCGCGAGCATCTGCGGCGGAGAGAGCTGACAAAGTGGAGAAATCCGC 60  
Db 3156 ATGGGCGCGCGCGCGAGGCTGTAGCGGCGGCGAGCTGGAACAAAGTGGAGAAATCCGC 3215  
Qy 61 CTGGCGCGCGCGCGCAAGAACATCATGCTGAAAGCACTGTGTGGGCGAGCGCGAG 120  
Db 3216 CTGGCGCGCGCGCGCAAGAAAGTACAGTGAAGCAATCGTGTGGGCGAGCGCGAG 3275  
Qy 121 CTGAAGGCTTGGCGCTGAACCCCGCGCTGTGAGACCGCGGAGGCTGCAAGCAGATC 180  
Db 3276 CTGAAGGCTTGGCGCTGAACCCCGCGCTGTGAGACCGCGGAGGCTGCGCAGATC 3335  
Qy 181 ATGAAGAGCTGACAGCGCGCGCTGCAAGCGGCAACGAGAGGCTGCGCTGTACAC 240  
Db 3336 CTGGGCGAGCTGACAGCGCGCGCTGCAAGCGGCAACGAGAGGCTGCGCTGTACAC 3395  
Qy 241 ACCGTGGCCACCTGTATCTGCTGCAAGCGCGGATCGAGGTCCGCAACCAAGAGGCC 300  
Db 3396 ACCGTGGCCACCTGTATCTGCTGCAAGCGGATCGAGGTCAAGAGCAACCAAGAGGCC 3455  
Qy 301 CTGAAGAAATCAAGAGAGGAGCAAGAACTCCCAAGAAAGCCAGAGCGCAAGAGAG 360  
Db 3456 CTGAAGAAATCAAGAGAGGAGAGAACTCCCAAGAAAGCCAGAGCGCGCGCGCC 3515  
Qy 361 GCCAGCGGCA-----AGGTAGCGCAAGACTAACCCCATGCGGCAAGACTG 405  
Db 3516 GCCAGCGGCAACCGGCAACAGACGAGGTAGCGCAAGACTAACCCCATGCGGCAAGACTG 3575  
Qy 406 CAGGCGCAAGTGTGACACAGGCGCATGAGCGCGGCAACCTGAAACGCTGGTGAAGTGT 465  
Db 3576 CAGGCGCAAGTGTGACACAGGCGCATGAGCGCGGCAACCTGAAACGCTGGTGAAGTGT 3635  
Qy 466 ATGAGAGAGAGGCTTCAAGCGCGCGGAGTATCCCAATGTTCAACCGCGCTGAGAGGCG 525  
Db 3636 GTGAGAGAGAGGCTTCAAGCGCGCGGAGTATCCCAATGTTCAACCGCGCTGAGAGGCG 3695  
Qy 526 GCCACCGCGCGAGACTGAAACAGATGTTGAACACCGTGGCGGCGACCAAGCGCGCATG 585



QY 646 CAGGCGGCGCCGCGCCGCGCGAGATGCGGACCCCGCGCGAGCGGACATCGCGCG 705  
DB 3893 CAGCGCGCGCCGCGCCGCGCGAGATGCGGACCCCGCGCGAGCGGACATCGCGCG 3952  
QY 706 GCCACGAGCACTCTGAGAGAGAGATGCGCTGATGACGAGCAACCCCGCGCGG 765  
DB 3953 ACCACGAGCACTCTGAGAGAGAGATGCGCTGATGACGAGCAACCCCGCGCGG 4012  
QY 766 GGGGACATCTACAAAGCGGTGATCATCTCTGGGCTTGAACAGATCGTGGATGACG 825  
DB 4013 GGGGAGATCTACAAAGCGGTGATCATCTCTGGGCTTGAACAGATCGTGGATGACG 4072  
QY 826 CCGGTGAGCATCTCTGAGCATCTCGCGAGGCGCCGAGAGGCGCTTCCGCGCATCGAG 885  
DB 4073 CCGACGAGCATCTCTGAGCATCTCGCGAGGCGCCGAGAGGCGCTTCCGCGCATCGAG 4132  
QY 886 CGCTTCTTCAAGACCTCTGCGCGCGAGAGAGGAGCGGAGCGTGAAGATGATGAC 945  
DB 4133 CGCTTCTTCAAGACCTCTGCGCGCGAGAGAGGAGCGGAGCGTGAAGATGATGAC 4192  
QY 946 GAGACCTGCTGAGAGAGAGCGGACACCCGAGTGAACCATCTCTGCGCGCTCTCGG 1005  
DB 4193 GAGACCTGCTGAGAGAGAGCGGACACCCGAGTGAACCATCTCTGCGCGCTCTCGG 4252  
QY 1006 CCGGCGCGCACTCTGAGAGAGATGATGACCGCTGCGAGGCGTGGGCGCGCGCGAC 1065  
DB 4253 CCGGCGCGCACTCTGAGAGAGATGATGACCGCTGCGAGGCGTGGGCGCGCGCGAC 4312  
QY 1066 AAGGCGCGCTGCTGCGCGAGGCGGATGAGCGAG---CGACAGCGTGAACATCATGAT 1122  
DB 4313 AAGGCGCGCTGCTGCGCGAGGCGGATGAGCGAGCGGCGGAGCGGCGGAGCGG 4372  
QY 1123 CAGAAAGCACTTCAAGGCGCGCGCGCGGAGGCGGAGGCGTGAAGTGGCGAG 1182  
DB 4373 CAGCGCGGCACTTCCGCGACAGCGGAGAGCGGAGGCGTGAAGTGGCGAG 4432  
QY 1183 GCGCACATGCGCAAGATCTGCGCGCGCGCGCGGAGAGGCGTGAAGTGGCGAG 1242  
DB 4433 GCGCACATGCGCAAGATCTGCGCGCGCGCGCGGAGAGGCGTGAAGTGGCGAG 4492  
QY 1243 GAGGCGCACTGAGAGAGATGAGATGAGCGGCGCGAGCGGCGGAGCGGAGGCG 1302  
DB 4493 GAGGCGCACTGAGAGAGATGAGATGAGCGGCGCGAGCGGCGGAGCGGAGGCG 4552  
QY 1303 CCGAGCGCAAGAGGCGCGCGCGGAGGCGGAGGCGGAGGCGGAGGCGGCGGCG 1362  
DB 4553 CCGAGCGCAAGAGGCGCGCGCGGAGGCGGAGGCGGAGGCGGAGGCGGCGGCG 4594  
QY 1363 ACCGTGCGCACTGCGCGCGCGCGCGGAGGCGGAGGCGGAGGCGGAGGCGGCG 1422  
DB 4595 CCGGAGCGCACTGCGCGCGCGCGCGGAGGAGGCGGAGGCGGAGGCGGAGGCG 4654  
QY 1423 AAGCAGAGGCGCAAGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 1482  
DB 4655 AAGCAGAGGCGCAAGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 4714  
QY 1483 TTGCGCAGGCGCGCGCTGAGCGGAGTAA 1509  
DB 4715 TTGCGCAGGCGCGCGGAGCGGAGGAGTAA 4741

## RESULT 6

US-09-475-515-7

Sequence 7, Application US/09475515A

Patent No. 6602705

GENERAL INFORMATION:

APPLICANT: BARNETT, Susan

APPLICANT: ZUR MEGEDE, Jan

APPLICANT: SRIVASTAVA, Indresh

APPLICANT: LIAN, Ying

APPLICANT: HARTOG, Karin

APPLICANT: LIU, Hong

APPLICANT: GREER, Catherine

APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475,515A  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7  
LENGTH: 2031  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-09-475-515-7

Query Match  
Best Local Similarity 88.5%; Fred. No. 5.6e-173;  
Matches 1349; Conservative 0; Mismatches 140; Indels 36; Gaps 3;

QY 1 ATGGCGCGCGCGCGAGCATCTGCGCGAGAGCTGAGCAAGTGGAGAGATCGCG 60  
DB 7 ATGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66  
QY 61 CTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 67 CTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126  
QY 121 CTGAGAGGCTTGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 127 CTGAGAGGCTTGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186  
QY 181 ATGAAGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 187 CTGAGAGGCTTGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
QY 241 ACCGTGCGCACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
DB 247 ACCGTGCGCACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
QY 301 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCGGAGCGCA-----AGTGAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 405  
DB 367 GCGGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426  
QY 406 CAGGCGCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465  
DB 427 CAGGCGCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486  
QY 466 ATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525  
DB 487 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546  
QY 526 GCGACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585  
DB 547 GCGACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606  
QY 586 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 645  
DB 607 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666  
QY 646 CAGGCGCGCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 705  
DB 667 CAGGCGCGCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
QY 706 GCGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765  
DB 727 ACCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786

QY 766 GGCGATCTACAGGCGTGGATCATCTTGGGCTGAACAAGATCGTGGATATGTAACG 825  
DB 787 GGCGAATCTACAGGCGTGGATCATCTTGGGCTGAACAAGATCGTGGATATGTAACG 846  
QY 826 CCCGTGAGCATCTTGAACATCCGCCAGGCGCCCAAGAGGCTTCCGCGCATGAGTAC 885  
DB 847 CCCACGAGCATCTTGAACATCCGCCAGGCGCCCAAGAGGCTTCCGCGCATGAGTAC 906  
QY 886 CGCTTCTTCAAGACCTTGGCGCGCGCGAGCAGGCGCAACCGAGACGTGAAGAACTGGATGAC 945  
DB 907 CGCTTCTTCAAGACCTTGGCGCGCGCGAGCAGGCGCAACCGAGACGTGAAGAACTGGATGAC 966  
QY 946 GAGACCTCTGTGTGAGAAAGCCCAACCCCGATGCAAGACCATCTTGGCGCGCTTCCGCG 1005  
DB 967 GAGACCTCTGTGTGAGAAAGCCCAACCCCGATGCAAGACCATCTTGGCGCGCTTCCGCG 1026  
QY 1006 CCCGCGCGCACCTTGAAGAGATGATGACCGCTTGGCGCGCGCGCGCGCGCGCGCGCG 1065  
DB 1027 CCCGCGCGCACCTTGAAGAGATGATGACCGCTTGGCGCGCGCGCGCGCGCGCGCGCG 1086  
QY 1066 AAGCGCGCTGTGTGAGAAAGCCCAACCCCGATGCAAGACCATCTTGGCGCGCTTCCGCG 1122  
DB 1087 AAGCGCGCTGTGTGAGAAAGCCCAACCCCGATGCAAGACCATCTTGGCGCGCTTCCGCG 1146  
QY 1123 CAGAGAGCAATCTTCAAGAGGCGCGCGCGCGCAACCTGATGCTTCAACTTGGCGCAAGAG 1182  
DB 1147 CAGCGCGCAATCTTCAAGAGGCGCGCGCGCGCAACCTGATGCTTCAACTTGGCGCAAGAG 1206  
QY 1183 GGCACATGCGCCAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242  
DB 1207 GGCACATGCGCCAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1266  
QY 1243 GAGGCGCACCAATGAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1302  
DB 1267 GAGGCGCACCAATGAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1326  
QY 1303 CCCAGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1362  
DB 1327 CCCAGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1368  
QY 1363 ACCGTGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1422  
DB 1369 CCCGAGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1428  
QY 1423 AAGCAGAGAGCCCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1482  
DB 1429 AAGCAGAGAGCCCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1488  
QY 1483 TTCGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1507  
DB 1489 TTCGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1513

RESULT 7  
US-09-475-515-5  
; Sequence 5, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30

NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1853  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-09-475-515-5  
Query Match 74.1%; Score 1117.8; DB 4; Length 1853;  
Best Local Similarity 85.1%; Pred. No. 16-160;  
Matches 1299; Conservative 0; Mismatches 192; Indels 36; Gaps 3;  
QY 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
DB 7 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66  
QY 61 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
DB 67 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 126  
QY 121 CTGGAGGCGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
DB 127 CTGGAGGCGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186  
QY 181 ATGAAGCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
DB 187 CTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246  
QY 241 ACCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
DB 247 ACCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306  
QY 301 CTGGAAGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGGAAGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCGGACGCGA-----AGGTGACCGAAGTACCCCATCTGTGAGAACTGT 405  
DB 367 GCGGACGCGA-----AGGTGACCGAAGTACCCCATCTGTGAGAACTGT 426  
QY 406 CAGGCGCAGATGTGACACAGGCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 465  
DB 427 CAGGCGCAGATGTGACACAGGCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486  
QY 466 ATCGAGAGAGAGGCGCTTCAAGCGCGCGAGGATCCCATGTTTCAAGCGCGCGCGCGCG 525  
DB 487 GTGAGAGAGAGGCGCTTCAAGCGCGCGAGGATCCCATGTTTCAAGCGCGCGCGCGCG 546  
QY 526 GCGACCGCGCGAGAGCTTGAACAGATGTTGAACAACGTTGAGCGCGCGCGCGCGCGCGCG 585  
DB 547 GCGACCGCGCGAGAGCTTGAACAGATGTTGAACAACGTTGAGCGCGCGCGCGCGCGCGCG 606  
QY 586 CAGATGCTGAAGAGACACATCAACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645  
DB 607 CAGATGCTGAAGAGACACATCAACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666  
QY 646 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 705  
DB 667 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726  
QY 706 GCGACGAGACCTTGAAGAGAGAGAGATGCGCTTGAAGACAGCAACCGCGCGCGCGCGCG 765  
DB 727 ACACACGAGACCTTGAAGAGAGAGATGCGCTTGAAGACAGCAACCGCGCGCGCGCGCG 786  
QY 766 GCGGACATCTACAGCGCGTGTATCTCTTGGCGCTTGAACAAGATGTGTGCGATGTACAGC 825  
DB 787 GCGGAGATCTACAGCGCGTGTATCTCTTGGCGCTTGAACAAGATGTGTGCGATGTACAGC 846  
QY 826 CCCGAGACATCTTGAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885

Db 847 CCCACGAGATCTCGAATCCGCGAGGGGCCCAAGAGGCTTCCGCGACTACGTGAC 906  
Qy 886 CGCTTTCTTCAAGACCTTGGCGCGCGAGCCGCAAGAGCTGAAAGCTGATGACC 945  
Db 907 CGCTTCTTCAAGACCTTGGCGCGCGAGCCGCAAGAGCTGAAAGCTGATGACC 966  
Qy 946 GAGACCTGTGTGTGAGAGGCAAGCCGCGAGTGAAGCAATCCGCGCGCTCTCGAC 1005  
Db 967 GAGACCTGTGTGTGAGAGGCAAGCCGCGAGTGAAGCAATCCGCGCGCTCTCGAC 1026  
Qy 1006 CCGCGCGCGCACTTGAAGAGATGATGACCGCTGCGAGGCGTGGGCGGCGCGCGAC 1065  
Db 1027 CCGCGCGCGCACTTGAAGAGATGATGACCGCTGCGAGGCGTGGGCGGCGCGCGAC 1086  
Qy 1066 AAGGCGCGGTGTGTGCGCGAGGCGAGTGAAGCAAGCTGAAAGCTGATGATG 1122  
Db 1087 AAGGCGCGGTGTGTGCGCGAGGCGAGTGAAGCAAGCTGAAAGCTGATGATG 1146  
Qy 1123 CAGAGAGCAATCTTCAAGAGGCGCGCGCGCAAGCTGCAAGTGTCTTCAAGTGGCGCAAGAG 1182  
Db 1147 CAGCGGCGCACTTCCGCAACGAGCGGAGAGCTGTAAGTGTCTTCAAGTGGCGCAAGAG 1206  
Qy 1183 GCGCAATGCGCAAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242  
Db 1207 GCGCAATGCGCAAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1266  
Qy 1243 GAGGCGCAATGAGTGAAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1302  
Db 1267 GAGGCGCAATGAGTGAAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1326  
Qy 1303 CCGAGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1362  
Db 1327 CCGAGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386  
Qy 1363 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1422  
Db 1369 CCGAGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1428  
Qy 1423 AAGCAGAGCGCGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1482  
Db 1429 TCTCAGAGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1488  
Qy 1483 TTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1509  
Db 1489 TTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1515

## RESULT 8

US-09-475-515-78  
Sequence 78, Application US/09475515A  
Patent No. 6602705  
GENERAL INFORMATION:  
APPLICANT: BARNETT, Susan  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: SRIVASTAVA, Indresh  
APPLICANT: LIAN, Ying  
APPLICANT: HARTOG, Karin  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475,515A  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 78  
LENGTH: 1865  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: GPI  
US-09-475-515-78

Query Match 74.1%; Score 1117.8; DB 4; Length 1865;  
Best Local Similarity 85.1%; Pred. No. 1e-160;  
Matches 1299; Conservative 0; Mismatches 192; Indels 36; Gaps 3;

Qy 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 13 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72  
Qy 61 CTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 73 CTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 132  
Qy 121 CTGGAGGCGCTTGCCTGTGAACCGCGCGCTGTGAGACCGCGCGAGGCGTGAACAGATC 180  
Db 133 CTGGAGGCGCTTGCCTGTGAACCGCGCGCTGTGAGACCGCGCGAGGCGTGAACAGATC 192  
Qy 181 ATGAAGCAGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 193 CTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 252  
Qy 241 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 253 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312  
Qy 301 CTGACCAAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360  
Db 313 CTGAGGAAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 372  
Qy 361 GCCGAGCGCA-----AGTGAAGCGCAATCTTCCGCGCGCGCGCGCGCGCGCGCG 405  
Db 373 GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 422  
Qy 406 CAGGCGCGAGTGTGACACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 465  
Db 433 CAGGCGCGAGTGTGACACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492  
Qy 466 ATCGAGAGGAAGGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 525  
Db 493 GTGAGGAGGAAGGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 552  
Qy 526 GCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 585  
Db 553 GCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612  
Qy 586 CAGATGCTGAAGAGCACCATCAACGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 645  
Db 613 CAGATGCTGAAGAGCACCATCAACGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 672  
Qy 646 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 705  
Db 673 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 722  
Qy 706 GCCACGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 765  
Db 733 ACCACGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 792  
Qy 766 GCGGACATCTTCAAGCGGCGGATCATCTGGCGCTGAACAAGATCGTGGGATGACAG 825  
Db 793 GCGGAGATCTTCAAGCGGCGGATCATCTGGCGCTGAACAAGATCGTGGGATGACAG 852  
Qy 826 CCGGTGAGCATCTTGAACATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885  
Db 853 CCGGCGCGCATCTTGAACATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 912  
Qy 886 CGCTTCTTCAAGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945  
Db 913 CGCTTCTTCAAGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 972  
Qy 946 GAGACCTGTGTGTGAGAGGCAAGCCGCGAGTGAAGCAATCCGCGCGCTCTCGAC 1005











RESULT 12  
US-09-952-060-27  
Sequence 27, Application US/09952060  
Patent No. 6733993  
GENERAL INFORMATION:  
APPLICANT: Emili A.  
APPLICANT: Emili, Emili A.  
APPLICANT: Yonil, Rima  
APPLICANT: Betc, Andrew J.  
APPLICANT: Chen, Ling  
APPLICANT: Kaslow, David C.  
APPLICANT: Shiver, John W.  
APPLICANT: Toner, Timothy J.  
APPLICANT: Casimiro, Danilo R.  
TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV-1 GAG, POL, NEF AND  
TITLE OF INVENTION: MODIFICATIONS  
FILE REFERENCE: 20747Y  
CURRENT APPLICATION NUMBER: US/09/952,060  
CURRENT FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: PCT/US01/28861  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/317,814  
PRIOR FILING DATE: 2001-09-07  
PRIOR APPLICATION NUMBER: 60/279,056  
PRIOR FILING DATE: 2001-03-27  
PRIOR APPLICATION NUMBER: 60/233,180  
PRIOR FILING DATE: 2000-09-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 27  
LENGTH: 1521  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Codon optimized DNA encoding human HIV-1 gag  
US-09-952-060-27

Query Match 64.4%; Score 971.4; DB 4; Length 1521;  
Best Local Similarity 79.6%; Pred. No. 1.4e-138;  
Matches 1210; Conservative 0; Mismatches 281; Indels 30; Gaps 4;  
QY 1 ATGGGCGCCGCGGAGCATCTCTGCGCGGAGGAGAGCTGAGCAAGTGGAGAGATCCGC 60  
DB 1 ATGGGCGCTAGGGCTTCTGTGCTGTCTGTGTGAGCTGAGCAAGTGGAGAGATCCGC 60  
QY 61 CTGGCCCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 61 CTGAGGCTGTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
QY 121 CTGAGAGGCTTGTGCGCTGTGAACCCCGGCTGTGTGAGAGAGAGAGAGAGAGAGAG 180  
DB 121 CTGAGAGAGGTTTGTGTGTGAACCTTGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 180  
QY 181 ATGAAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 181 CTGGGCGAGCTTCAAGCTCTCTGTGCAACAGGCTCTGAGAGAGAGAGAGAGAGAG 240  
QY 241 ACCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
DB 241 ACAGTGTGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300  
QY 301 CTGAGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 353  
DB 301 CTGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 354 --CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 411  
DB 354 GAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
QY 412 CAGATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471  
DB 421 CAGATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480

QY 472 GAGAAAGCCTTTCAGCCCGAGAGTGTATCCCATTTTCAACCGCCTGAGAGAGAGAGAG 531  
DB 481 GAGAAAGCCTTTCAGCCCGAGAGTGTATCCCATTTTCTGCGCCTGTGTGAGAGAGAGAG 540  
QY 533 CCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 591  
DB 541 CCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
QY 592 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651  
DB 601 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 652 GGGCCCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 711  
DB 661 GGGCCCGATTTGGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 712 AGCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771  
DB 721 TCACCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
QY 772 ATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 831  
DB 781 ATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
QY 832 AGCATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
DB 841 TCACCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 892 TTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951  
DB 901 TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 952 CTGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
DB 961 CTGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
QY 1012 GCGACCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071  
DB 1021 GCGACCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1072 GCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128  
DB 1081 AGGAGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1129 AGCAACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188  
DB 1141 GCGCAACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
QY 1189 ATGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1248  
DB 1201 ATGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1249 CACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308  
DB 1261 CACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
QY 1309 CACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368  
DB 1321 CACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
QY 1369 CCCAGCGCCCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1428  
DB 1381 GAGTCTTTC-----AGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1428  
QY 1429 GAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488  
DB 1429 GAGCCCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482  
QY 1489 AGCGGCGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1509  
DB 1483 AACGAGCCCTTCTCCAGTAA 1503

RESULT 13  
US-09-974-702-1  
Sequence 1, Application US/09974702  
Patent No. 6696291  
GENERAL INFORMATION:  
APPLICANT: Shiver, John W.  
Davies, Mary Ellen  
Freed, Daniel C.  
Lin, Margaret A.  
Perry, Helen C.  
TITLE OF INVENTION: Synthetic HIV Gag Genes  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merck & Co., Inc.  
STREET: P.O. Box 2000, 126 E. Lincoln Ave.  
CITY: Rahway  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/974,702  
FILING DATE: 09-Oct-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/017,981  
FILING DATE: 03-FEB-1998  
APPLICATION NUMBER: US60/037,854  
FILING DATE: 07-FEB-1997  
APPLICATION NUMBER: GB9705040.5  
FILING DATE: 12-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Hand, J. Mark  
REGISTRATION NUMBER: 36,545  
REFERENCE/DOCKET NUMBER: 19730  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732-594-3905  
TELEFAX: 732-594-4720  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1532 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-974-702-1  
Query Match 64.4%; Score 971.4; DB 4; Length 1532;  
Best Local Similarity 79.6%; Pred. No. 1.4e-138;  
Matches 1210; Conservative 0; Mismatches 281; Indels 30; Gaps 4;

Db 190 CTGGGCGAGCTCCAGCCCTCCCTGCAACAGGCTCGAGGAGCTGAGTCCCTGTACAAAC 249  
QY 241 ACCGTGGCCACCTCTGATCTGCTGCAAGCCGCGCATGAGTCCGCGACACCAAGAGGCC 300  
Db 250 ACAGTGGCTACCTCTGATCTGCTGCAAGGAGTGTGTAAGAGACCAAGAGAGGCC 309  
QY 301 CTGGAAGAATGAGAGGAGAGAGAAAGTCCCGAGAGAACCCAGAGGCC----- 353  
Db 310 CTGGAAGAATGAGAGGAGAGAGAAAGTCCAGAGAGAGGCCCGAGAGGCTGCTGCT 369  
QY 354 --CAAGAGGCCGAGAGGAGAGAGTGAAGCCAGAACTACCCCATGTCGAGAACTGCAAGGCC 411  
Db 370 GGCACAGGAACTTCCAGCCAGGCTGTCAGAACTACCCCATGTCGAGAACTCCAGAGGCC 429  
QY 412 CAGATGTGACACAGGCGCATGAGCCCGCGACCTTGAAAGCTTGAGTGAAGTATGAG 471  
Db 430 CAGATGTGACACAGGCGCATGAGCCCGCGACCTTGAAAGCTTGAGTGAAGTATGAG 489  
QY 472 GAGAGGCTTCAAGCCCGAGAGTGAATCCCATGTTCAAGCCCTGAGAGAGGCGCGACCC 531  
Db 490 GAGAGGCTTCTCCCTGAGGTGAATCCCATGTTCTGAGGCTGTCGAGGAGTGCACCC 549  
QY 532 CCCAGAGCTTGAACAGATGTTGAACACCGTGGCGCGCACAGGCGCCCATGAGATG 591  
Db 550 CCCAGAGCTTGAACAGATGTTGAACACAGTGGGGGGGCATCAGGCTGCATGAGATG 609  
QY 592 CTGAAGAGACCATCAACAGAGAGGCCCGGAGTGGAGCCGCTGACACCCCGTCAAGGCC 651  
Db 610 CTGAAGAGAGCATCAATGAGAGGCTGTGAGTGGGACAGGCTCATCTCTGTGACCGCT 669  
QY 652 GGGCCCGTGGCGCCGCGAGATGGCGACCCCGCGGAGAGATGCGCGGCGCGACCC 711  
Db 670 GGGCCCGTGGCGCCGCGAGATGGCGAGCCCGAGGCTGTGACATGCTGTGACACACC 729  
QY 712 AGCACCTTGAAGAGAGAGTGCCTGATGACAGACACCCCGCGTGGCGCGAGC 771  
Db 730 TCCACCTTGAAGAGAGAGTGGCTGATGACACACACCCCGCATCTGTGGGGGAGAA 789  
QY 772 ATCTCAAGAGGCTGATCATCTTGGGCTGAAACAGATGCTGCGATGTACAGCCCGTGG 831  
Db 790 ATCTCAAGAGGCTGATCATCTTGGGCTGAAACAGATGCTGAGATGTACCTCCCGCAC 849  
QY 832 AGCATCTGAGACATCCCGAGAGGCCCGCAAGAGCCCTCCGAGCATGCTGAGACCGCTTC 891  
Db 850 TCCATCTGAGACATGAGGAGGCCCGCAAGAGCCCTTCAGGAGCTATGTGACAGGTTC 909  
QY 892 TTCAAGACCTTGGCGCGAGAGAGGCCCAAGAGCTGAAAGTGAATGATGACCGAGACC 951  
Db 910 TACAAAGACCTTGAAGAGGCTGAGAGAGGCTTCCAGAGAGTGAAGAACTGATGACAGAGAC 969  
QY 952 CTGCTGTGAGAGAGAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
Db 970 CTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1029  
QY 1012 GCCACCTTGAAGAGAGATGATGACCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071  
Db 1030 GGCACCTTGAAGAGAGATGATGACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1089  
QY 1072 CGCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128  
Db 1090 AGGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1149  
QY 1129 AGCAACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188  
Db 1150 GGCACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1209  
QY 1189 ATGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1248  
Db 1210 ATTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1269  
QY 1249 CACGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308

Db 1270 CACCAATGATGAAGCATGTCATGAGAGGAGGCACTTCTGGGCAAAATCTGGCCCTCC 1329  
Qy 1309 CACAAAGGCGCGCCCGGCAATTCTTGAGAACCGAGCGAGCCGCGCCGCCACCCCTG 1368  
Db 1330 CACAAAGGCGCGCGGCACTTCTTGAGAACCGAGCGAGCCGCGCCGCCACCCCTG 1389  
Qy 1369 CCGACGCGCGCGCGCGGAGAGCTTCCGTTGAGAGAGACACCCCGCGCCGAGAG 1428  
Db 1390 GAGTCTCTT-----AGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
Qy 1429 GAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488  
Db 1438 GAGCCCATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1491  
Qy 1489 AGCGGCGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1509  
Db 1492 AAGGAGCGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1512

RESULT 14  
US-09-818-443-1  
Sequence 1, Application US/09818443  
Patent No. 6787351  
GENERAL INFORMATION:  
APPLICANT: Chen, Ling  
APPLICANT: Shiver, John W.  
APPLICANT: Betz, Andrew J.  
APPLICANT: Casimiro, Danilo R.  
APPLICANT: Caulfield, Michael J.  
APPLICANT: Chastain, Michael A.  
APPLICANT: Emili, Emilio A.  
TITLE OF INVENTION: ADENOVIRUS CARRYING GAG GENE HIV VACCINE  
FILE REFERENCE: 20440YIA  
CURRENT APPLICATION NUMBER: US/09/818,443  
CURRENT FILING DATE: 2001-09-13  
PRIOR APPLICATION NUMBER: PCT/US00/18332  
PRIOR FILING DATE: 2000-07-03  
PRIOR APPLICATION NUMBER: 60/148,981  
PRIOR FILING DATE: 1999-08-13  
PRIOR APPLICATION NUMBER: 60/142,631  
PRIOR FILING DATE: 1999-07-06  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FaastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1532  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Optimized human HIV-1 gag ORF  
US-09-818-443-1

Query Match 64.4%; Score 971.4; DB 4; Length 1532;  
Best Local Similarity 79.6%; Pred. No. 1.4e-138;  
Matches 1210; Conservative 0; Mismatches 281; Indels 30; Gaps 4;  
Qy 1 ATGGGCGCGCGCGCGGAGATCTTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
Db 10 ATGGGTGCGAGGGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 69  
Qy 61 CTGGCGCGCGCGCGGAGAGAGATCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
Db 70 CTGAGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 129  
Qy 121 CTGAGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180  
Db 130 CTGAGAGAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 189  
Qy 181 ATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 190 CTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 249  
Qy 241 ACCGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

Db 250 AAGTGGCTAACCTGTATCTGTGTGACCAAGAGATGTGTGAAGAGAGAGAGAGAG 309  
Qy 301 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 353  
Db 310 CTGGAAGAGATGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369  
Qy 354 --CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 411  
Db 370 GGCACAGAGCAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429  
Qy 412 CAGATGTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471  
Db 430 CAGATGTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489  
Qy 472 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531  
Db 490 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 549  
Qy 532 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 591  
Db 550 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609  
Qy 592 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651  
Db 610 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 669  
Qy 652 GAGCGCGTGTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 711  
Db 670 GAGCGCGTGTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729  
Qy 712 AGCAGCGTGTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771  
Db 730 TCACCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 789  
Qy 772 ATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 831  
Db 790 ATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 849  
Qy 832 AGCATCTGTGAGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
Db 850 TCATCTCTGTGAGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 909  
Qy 892 TTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951  
Db 910 TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 969  
Qy 952 CTGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
Db 970 CTGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1029  
Qy 1012 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071  
Db 1030 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1089  
Qy 1072 GCGGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128  
Db 1090 AGGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1149  
Qy 1129 AGCAACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188  
Db 1150 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1209  
Qy 1189 ATGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1248  
Db 1210 ATTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1269  
Qy 1249 CACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308  
Db 1270 CACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1329  
Qy 1309 CACAAAGGCGCGCGCGGCAATTCTTGAGAACCGAGCGAGCCGCGCCGCCACCCCTG 1368  
Db 1330 CACAAAGGCGCGCGGCAATTCTTGAGAACCGAGCGAGCCGCGCCGCCACCCCTG 1389



```
QY      1369  CCCACGCCCCCCCCCGAGAGCTTCCGCTTCAGAGAGACCAACCCCGCCCCCAAGCAG 1428
          |||
          2652  GAGTCTTC-----AGTTTGGGAGAGAGAGACCAACCCCGAGCAG 2699
          |||

QY      1429  GAGCCCAAGAGACCGGAGCCCTACCGCGAGCCCTGACCGCCCTGCGCAGCCTGTTGGGC 1488
          |||||
          2700  GAGCCCATTTGACAAGAGAGCTGAC-----CCCTGGCCTCCCTGAGGTCCCTGTTGGC 2753
          |||||

QY      1489  AGCGGCCCCCTGAGCCAGTAA 1509
          |||
          2754  AACGACCCCTCTCCCAGTAA 2774
          |||||
```

Search completed: March 13, 2005, 00:22:49  
Job time : 273.651 sec9









Db 617 GGAATTCCTCAGGCGCCATGAAGATCGAGATGCCACCTTTCCTTCATCAGGACGG 676  
 QY 846 CCGCAGAGGCGCCCAAGAGCCCTTCCGCACTACGTGACCCGCTTCTTCAAGACCTCGC 905  
 Db 677 CCAAGATGTCGGCGCGCTACGCTCCGCGCAAGGGGAGAGCTGTCGTGTGATCTCTCCG 726

RESULT 4  
 CN128942 826 bp mRNA linear EST 01-APR-2004  
 LOCUS RHOH1\_32\_C05\_g1\_A002 Acid- and alkaline-treated roots Sorghum  
 DEFINITION bicolor cDNA clone RHOH1\_32\_C05\_A002 5', mRNA sequence.  
 ACCESSION CN128942  
 VERSION CN128942.1 GI:45956257  
 KEYWORDS EST.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 BUKARYOTA; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 826)  
 Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C.,  
 Sun, F., Sullivan, R., Lim, S., Eastman, A. and Pratt, L.H.  
 An EST database from Sorghum: acid- and alkaline-treated roots  
 Unpublished (2003)  
 Other ESTs: RHOH1\_32\_C05\_b1\_A002  
 Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmp@pratt@uga.edu

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
 the Human Genome Center, University of Tokyo Institute of Medical  
 Science; plant material and RNA prepared at Texas A & M University;  
 sequencing done in the Laboratory for Genomics and Bioinformatics,  
 University of Georgia. Sequence ends have been trimmed to exclude  
 vector and regions below phred quality 16. Three-prime sequences  
 are presented as their reverse complement and have been trimmed to  
 exclude polyA.  
 Seq primer: Sugs (CTTCTGCTTAAAGCTCG).  
 Location/Qualifiers

FEATURES  
 source  
 1..826  
 /organism="Sorghum bicolor"  
 /mol\_type="mRNA"  
 /cultivar="Brix623"  
 /db\_xref="taxon:4558"  
 /clone="RHOH1\_32\_C05\_A002"  
 /lab\_host="DH10B-T1 phage-resistant E. coli"  
 /clone\_lib="Acid- and alkaline-treated roots"  
 /note="Organ: Root; Vector: pME18S-FL3; Site 1: XhoI;  
 Site 2: XhoI; The library was prepared from polyA+ RNA  
 from 8-day-old roots harvested from Brix623 sorghum  
 seedlings grown in hydroponic culture. HCl was added to a  
 pH of 3.0 to some seedlings, KOH to a pH of 9.0 for  
 others. Roots were harvested 3, 12 and 27 hr after  
 addition and pooled for RNA preparation. Double-stranded  
 cDNA was cloned unidirectionally into different DraIII  
 sites of the pME18S-FL3 vector (5-prime DraIII site is  
 CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises  
 the cDNA insert."

## ORIGIN

Query Match 5.4%; Score 81.6; DB 7; Length 826;  
 Best Local Similarity 47.0%; Freq. No. 9.9e-05;  
 Matches 358; Conservative 0; Mismatches 394; Indels 10; Gaps 3;  
 QY 2 TGGGCGCCCGCCGACGATCTGCGCGCGAGAGCTGCAAGTGGAGAGATCCGCC 61  
 Db 60 TCCCGCGATGCGACACCGAGCGGCGCGCCGCGCGCGAGCGGCGCA 119  
 QY 62 TCGCGCCCGCGCGCAAGAACCTACATCTGAAGCACTGTGTGTGGCGCAGCCGAGC 121

Db 120 ACGGCAATGCGCAACGCGGAGCAAGAGCGCCACTTCGAGGTCCGACAACAAGACCTTGC 179  
 QY 122 TGAAGGAGCTTTCGCTTGAACCCCGGCTGTGTGAGACCGCCGAGGGCTGCAAGCATCA 181  
 Db 180 TCAAGAGCGACGACCTCTACGAGTACATCTTGAACAGAGGTGTACCCGGGGAGCCGG 239  
 QY 182 TGAAGCACTGACGCGCCGCTTGCAGACCGGAGCCGAGAGAGTGGCGAGCTGTACAACA 241  
 Db 240 AGAGCATGAAGAGGTCCGCGAGATCACCGCAGACCACTGAACTGATGAGCACT 299  
 QY 242 CCGTGGCCACCTGTACTGCGGTGCAAGCGCGGATGAGTCCGCAACCAAGAGGCC 301  
 Db 300 CCGCGGACGAGGGGAGTTCCTCAACATGCTATCAACTCATGCGCGCAAGAGCA 359  
 QY 302 TGAAGCAATGAGAGAGAGAGAGCAAACTCCAGACAGAACCCAGAGGCGCAAGAGAG 361  
 Db 360 TGAATCATCAACCGGAGAACTACAGAGCTGGGCTGCTCCATGCAAGAAAGCCGCGCTCG 419  
 QY 362 CCGAGCGCAAGGTGAGCCGAACTACCCCATCTGCAAGAACTGCAAGGCGCCAGATGTGC 421  
 Db 420 CCCAGAGATGCACTTCCGCGAGGCGCCCGGCTCCGCTCCGACGACCTCATGCGCG 479  
 QY 422 ACCAGGCATCAGCCCGGACCCCTGAAGCGCTGGTGAAGTATGAGAGAGAGGCT 481  
 Db 480 ACGAAGAAACGACGAGGTCTGCTTCACTTCTTGTGAGAGCCGACAGAAACAACTAC 539  
 QY 482 TGAAGCCGAGGTATCCCATGTTTACCGCCCTGAGAGAGGCGCAACCCCGAGACC 541  
 Db 540 TCAACTTACACG---ACGGGCTGCTCAAGTGTGTAAGTGGGGGCTCATCGGCTACG 596  
 QY 542 TGAACAGATGTGTAACACCGTGGGCGCCACAGAGCCGCGCATGTGCTGAAGACA 601  
 Db 597 ACAACACGCTGTGAACGCGCTCTGTGCTCCCGAGAGAGCCCGCATGTGCAAGTACA 656  
 QY 602 CCATCAAGAGAGAGCGCGCGAGTGGAGCCGCTGCAACCGTGAAGGCGGCGCCGCTGG 661  
 Db 657 TCCGTTTCA-----CGGCACTTCTGCTGTCTTCTCAAGAGCGTCCGCGCCAGC 710  
 QY 662 CCGCGGCGCAATGCGGACCCCGCGGAGCGAGCATGCGCGGCGCCACAGCAACCTTGC 721  
 Db 711 AGCGGTGATGATGTGCGAGCTCCCGCGGAGAGGTACACCTTGTGCGC-CGCGTC 769  
 QY 722 AGAGCAGATGCTGTGATGACAGCAACCCCGCGTGGCG 763  
 Db 770 AAGTAAAAAACAATGCTGCGCACCGAGCGCGCGCGCG 811

RESULT 5  
 LOCUS CD500131/C 1219 bp mRNA linear EST 12-JUN-2003  
 DEFINITION CD443-C08\_y1d-g SHGC-CD4 Gasterosteus aculeatus cDNA clone  
 CD500131  
 ACCESSION CD500131  
 VERSION CD500131.1 GI:11427162  
 KEYWORDS EST.  
 SOURCE Gasterosteus aculeatus (three spined stickleback)  
 ORGANISM Gasterosteus aculeatus  
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;  
 Gasterosteidae; Gasterosteus.  
 1 (bases 1 to 1219)

REFERENCE  
 AUTHORS Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,  
 Schmutz, J., and Myers, R.M.  
 TITLE Expressed sequence tags from Gasterosteus aculeatus  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Kingsley, DM  
 HHMI and Department of Developmental Biology  
 Stanford University School of Medicine  
 Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA  
 Tel: 650 725 5954  
 Fax: 650 725 7739

Email: Kingsley@cimgm.stanford.edu

Plate: 43

High quality sequence start: 4

High quality sequence stop: 792.

## FEATURES

Source

1. 1219

/organism="Gasterosteus aculeatus"

/mol\_type="mRNA"

/strain="Salinas river, CA"

/db\_xref="taxon:69293"

/clone="CDA43-C08"

/sex="mixed male and female"

/tissue\_type="heads and internal organs combined"

/dev\_stage="adult"

/clone\_lib="SHG-CDA"

/note="Vector: lambda ZAP Express/PBK-CMV, Site 1: EcoRI (5' adaptor); Site 2: XhoI (3' linker primer); The mixed organ cDNA library was generated using the ZAP-cDNA method by Stratagene. First strand cDNA synthesis was primed with a 50 bp linker primer containing an oligo dt sequence preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector undirectionally in the sense orientation with respect to the lacZ promoter of PBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual PBK-CMV phagemid clones for EST sequencing."

## ORIGIN

Query Match

Best Local Similarity 47.7%; Score 81.6; DB 6; Length 1219;

Matches 373; Conservative 0; Mismatches 394; Indels 15; Gaps 4;

```

QY 618 CGCCGAGTGGAGACCGGCTGACACCGGTCAGGCGGCGCCGTCGCGCCGCGCAGATGGG 677
DB 917 CGGCTCGCTGGGCGACTACGGCGGGGTAACGGCGACGCGCTCTCAACGGCTCGCTGGG 858
QY 678 CGACCCCGCGGCGAGGAC-----ATGCGCGGCGGCACACGAGCCTTCAAGAGAGAT 731
DB 857 CGACTACGCGGGGCGAGCGCCCTGCTCAACGGCTCGTGGGCGACTACGGCTACGGCGCA 798
QY 732 CGCCTGATGACCAACACCCCGCTGCGCCGTCGCGGCGACATTAACCGGTGATCAT 791
DB 797 CGCCTGCTCAACGGCTCGCTGGGCGACTACGGCTACGGCGGCGCCTCTCAACGGCTC 738
QY 792 CCGGCGCTGAGCAAGATGTCGATGTACAGCCCGGAGCATCTGAGACATCCGCGCA 851
DB 737 GCTGGGCGACTACGGCGAGCGCTGCTCAACGGCGAGCTGGGCGACTACGGCGCGCGCA 678
QY 852 GGGC---CCCAAGAGACCCCTTCGCGACTACGTGACCGCTTCTTCAAGACCTGTC---G 905
DB 677 CGCCTGCTCAACGGCTCGCTGGGCGACTACGGCGGGGTAACGGCGAGCGCTGCTCA 618
QY 906 CGCCGAGCGAGCCACCAAGAGCTGAAGAACTGATGACCGAGACCTGCTGTGTGAGAA 965
DB 617 CGGCTCGCTGGGCGACTACGGCGGCGAGCGCCCTGCTCAACGGCTCGCTGGGCGACTA 558
QY 966 CGCGAACCCCGACTGCAAGACATCTGGCGGCTCGCGCCCGGCGGCGACCTGAGAGA 1025
DB 557 CGGCTACGGCGAGCGCTGCTCAAC--GACTGCTGGGCGAGTACGGCTACGGCGCGCG 501
QY 1026 GATGATGACCGCTGTCAGAGCGCTGGGCGGCGCCCGGCGCAACAGCCCGCTGTCGCG 1085
DB 500 CCGTCTCAACGGCTCGCTGGGCGACTACGGCGAGCGCTGCTCAACGGCGAGCGCTGGGCG 441
QY 1086 GGGGATGAGCCAGGCGCAACAGCTGAACATGATGACAGAGAGCACTTCAAGGCGCC 1145
DB 440 CTACAGCGCGGCGAGCGCTGCTCAACGGCTGCTGGGCGAGTACGGCTACGGCGAGCG 381
QY 1146 CCGGCGCAACGTGCTTCAACTGCGGCGAGAGGCGCAATCGCCCAAGACTGCGCG 1205

```

```

DB 380 CCGTCTCAACGGCTGCTGGGCGAGTACGGCGGCGGCTACGGCGGCGCCCTGCTCAACGG 321
QY 1206 CGCCCCCGCAAGAGGCTGCTGGAAGTGGCGGCAAGAGGCGCACCAATGAAGACTG 1265
DB 320 CTCGCTGGGCGACTACGGCGGCGGCTACGGCGCGCCCTGCTCAACGGCGAGCTGGGCGCA 261
QY 1266 CACCGAGGCGCGAGCGCACTTCTGCGGCAAGATCTGGGCCAGCGCAAGAGGCGCGCCGG 1325
DB 260 CTACGCGCGCTACGGCGAGCGCTGCTCAACGGCGAGCTGGGCGAGTACGGCGGCTACGG 201
QY 1326 CACTTCTGCAAGAACCGGAGGAGCGCGCGCCCGCACTGTCACCGCGCCCGCCCGCG 1385
DB 200 CGACCGCTGCTCAACGGCTGCTGGGCGAGTACGGCGGCTACGGCGAGCGCCCTGCTCA 141
QY 1386 CG 1387
DB 140 CG 139

RESULT 6
CL965300 1800 bp DNA linear GSS 21-SEP-2004
LOCUS OSIFC012134 Oryza sativa Express Library Oryza sativa (Indica
DEFINITION cultivar-group) genomic, genomic survey sequence.
ACCESSION CL965300.1 GI:52385293
VERSION CL965300
KEYWORDS GSS.
SOURCE Oryza sativa (Indica cultivar-group)
ORGANISM Oryza sativa (Indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriactroideae; Oryzaceae; Oryza.
1 (bases 1 to 1800)
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G. K. S., Deng, X. W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatics
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-8048676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
Location/Qualifiers
1..1800
/organism="Oryza sativa (Indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences"

ORIGIN
Query Match 5.4%; Score 81.6; DB 9; Length 1800;
Best Local Similarity 44.2%; Pred. No. 0.0001;
Matches 492; Conservative 0; Mismatches 599; Indels 21; Gaps 3;

```



QY 913 CAGGCCACCCAGAGCTGAAGAACTGATGACCGAGACCTCTGCTGACGAAGCCCAAC 972  
 DB 868 CACGGCCGGACATCAACCGCGCGCGCGCCCTGGGCAAGCTCCGCGCGAGTCCGAG 927  
 QY 973 CCGGATGGAAGACATCTCTGCGGCTTGGGCGCCCGGCGCAACCTTGAAGAGATGATG 1032  
 DB 928 CCGG---CCAGGCGCGGCTCAGCAACACAGACAGAGTGGCGCTCAGAGTCCGCTG 984  
 QY 1033 ACCGCTGCGCAGGCGCTGGGCGCGCCCGCGCAAGAGCCCGCTGTGCGCGAGGCGATG 1092  
 DB 985 TTCGACGGCGTGCATCTTGTCCGAGCGGCTCTCCCGGCGCGGTTCCAGAGGCTCAACAG 1044  
 QY 1093 AGCCAGCCCAACAGCGCTGACATCATGATGACAGAAAGCACTTCAAGGCGCCCGCGC 1152  
 DB 1045 GACCTCTTCAAGAAAGACGATGGTGGCGGTGAGGAGGAGCCATGCGGACCGCCGCTGAGC 1104  
 QY 1153 AAGCTCAAGTGTCTTCAACTGCGGCAAGAGGCGCATATGCGCCCAAGACTGCGCGCCCG 1212  
 DB 1105 AAGGCGCATCATGACGACGATCTCTCTGCGCGGCGACACCGAGATCCCAAGGTGACG 1164  
 QY 1213 CCGAAGAGGCGCTGCTGAAGTCCGCGCAAGAGGCGCACCAATGAAGACTGACCGAG 1272  
 DB 1165 CAGCTGCTCAAGGACTACTTCTGGTGGCAAGAGCCCAACCGCGCGTCAACCCCGAGAG 1224  
 QY 1273 CCGGAGCCCACTTCTCTGGGCAAGATCTGGCCCAAGCAAGGCGCCCGCGCACTTG 1332  
 DB 1225 GCGGTGGCTACGCGCGCGCGCGCGCGAGCGCATATCATGCGCGCACCTTCAACGAGAA 1284  
 QY 1333 CTGCAAGACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 1392  
 DB 1285 ACGGAGAGATGATCTCTCTGACGCTGCGCGCGCTCAACCTCGGCTTGAAGAGCGCGCG 1344  
 QY 1393 TTCCGCTTGAAGAGACCAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1439  
 DB 1345 GCGGTGATGCGCAAGCTGATCCCGCGCAACAGCGTGTGTCGACGAA 1391

RESULT 8  
 BQ744212 749 bp mRNA linear EST 17-JUL-2002  
 LOCUS WHE4113\_A01 A01ZS Wheat salt-stressed root cDNA library Triticum  
 DEFINITION aestivum cDNA clone WHE4113\_A01\_A01, mRNA sequence.  
 ACCESSION BQ744212  
 VERSION BQ744212.1 GI:21890999  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooideae; Triticaceae; Triticum.  
 1 (bases 1 to 749)  
 Dvorak, J., Lazo, G.R., Pham, J., Rausch, C.J., Wilson, C. and Woo, J.  
 Anderson, O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K.,  
 The structure and function of the expressed portion of the wheat  
 genome - Salt-stressed root cDNA library  
 Unpublished (2002)  
 JOURNAL Contact: Olin Anderson  
 COMMENT US Department of Agriculture, Agriculture Research Service, Pacific  
 West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 5105858773  
 Fax: 510585818  
 Email: oanderson@pw.ars.gov  
 Sequences have been trimmed to remove vector sequence and low  
 quality sequence with phred score less than 20  
 Seq primer: SK primer.  
 FEATURES  
 Source Location/Qualifiers  
 1..749  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"

/clone="WHE4113\_A01\_A01"  
 /tissue\_type="Roots"  
 /dev\_stage="Full tillering"  
 /lab\_host="E. coli SOLR"  
 /clone\_1ib="Wheat salt-stressed root cDNA library"  
 /note="Vector: lambda uni-zap X, excised phagemid  
 plasmid SK(-); Site 1: EcoRI; Site 2: XhoI; Hydroponic  
 plants grown to full tillering stage were treated with 150  
 mM NaCl for either 12 hours or 7 days. Root tissues of the  
 plants subjected to both types of treatment were collected  
 separately at University of California, Davis (E. Akhunov  
 and K. Deal in J. Dvorak's lab). Total RNA was prepared  
 separately from the two samples (12h and 7day treatments),  
 and equal amount of RNA was then pooled. PolyA RNA was  
 purified from the pooled RNA, a cDNA library was made, and  
 the cDNA clones were in vivo excised to give plasmid  
 SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak)  
 at the University of California, Davis. Colony plating,  
 plasmid DNA preparations and DNA sequencing were performed  
 in the OD Anderson lab (all other authors)."

## ORIGIN

Query Match 5.4%; Score 81; DB 5; Length 749;  
 Best Local Similarity 48.4%; Pred. No. 0.00013;  
 Matches 293; Conservative 0; Mismatches 300; Indels 12; Gaps 2;  
 QY 269 CCGGATGAGGTCCGCGACACCAAGAGGCGCTTGAAGATCGAGAGAGAGAGACA 328  
 DB 8 CTTCAAGCTCTTCACACGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67  
 QY 329 AGTCCAGAGAAAGCCCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 388  
 DB 68 GCTTCAACCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 127  
 QY 389 CCATCGTGAAGAACTGCAAGGCGCGAGATGTCACACGAGCGCATACGCGCGCGCGCG 448  
 DB 128 GCGAGCGCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 187  
 QY 449 ACGCTGGG--TGAAGTGAATGAGAGAGGCGCTTCAAGCCCGAGGTATCCCATGT 505  
 DB 188 ACCACAGAGGCTTCCGCGCGCAACAGAGAAACCGCGCAAGCGCTTCCGATCAGAC 247  
 QY 506 TCAAGCGCTGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 565  
 DB 248 CCGACTGACCATCTCTGACGAGATCCCTTGGCAACTTCAAGCTCTCTTGGCG 307  
 QY 566 GCGGCAACAGGCGCGCATGACATGCAAGACCATCAACGAGAGGCGCGCGAGT 625  
 DB 308 TTGCGGACCAAGCGCGAGACCTGCTGCTGTGGGCGCGTCACTCTGACACCGCGCT 367  
 QY 626 GGGACCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 685  
 DB 368 ACGACCGGTCACACCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 427  
 QY 686 GCGGAGGACATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 736  
 DB 428 TCAAGATCAACCAACGACGACCGCATATCTGCGCGCTTCCGAGAGGAGCAAGGCCA 487  
 QY 737 GGATGACCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 796  
 DB 488 CCGTCTTGGACCAAGCGCGCATCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547  
 QY 797 GCTTGAACAAAGTGTGGGATGTACAGCGCGCGTGAACATCTGACATCCGCCAGGCG 856  
 DB 548 CTTGAGATATGTGTGAGCGCGCGTGGCAACAACTCTTCTTGAAGAGCGCGCGCGCT 607  
 QY 857 CCAAG 861  
 DB 608 CCGAG 612  
 RESULT 9  
 BE636712

LOCUS BE636712 1674 bp mRNA linear EST 03-JAN-2002  
 DEFINITION rockefeller.0.42 Masticamoeba balamuthi lambda ZAP II library  
 Masticamoeba balamuthi cDNA similar to vacuolar ATP synthase  
 subunit B (EC 3.6.1.3), mRNA sequence.  
 ACCESSION BE636712  
 VERSION BE636712.1 GI:9919823  
 KEYWORDS EST.  
 SOURCE Masticamoeba balamuthi  
 ORGANISM Masticamoeba balamuthi  
 Eukaryota; Pelobiontida; Masticamoebidae; Masticamoeba.  
 REFERENCE 1 (bases 1 to 1674)  
 Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W.,  
 Gordon, P., Dunfield, L., Gaasterland, T., Lopez, P., Muller, M. and  
 Philippe, H.: The analysis of 100 genes supports the grouping of three highly  
 divergent amoebae: Dictyostelium, Entamoeba, and Masticamoeba  
 Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
 MEDLINE 21819461  
 PUBMED 11830664  
 COMMENT Contact: Muller Miklos  
 Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockefeller.edu  
 Insert Length: 1674 Std Error: 0.00  
 POLY=A=yes.  
 FEATURES  
 source Location/Qualifiers  
 1..1674  
 /organism="Masticamoeba balamuthi"  
 /mol\_type="mRNA"  
 /strain="ATCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_id="Masticamoeba balamuthi lambda ZAP II library"  
 /note="syn: Phreatamoeba balamuthi"  
 ORIGIN  
 Query Match 5.3%; Score 80.6; DB 2; Length 1674;  
 Best Local Similarity 44.1%; Pred. No. 0.00015;  
 Matches 615; Conservative 0; Mismatches 756; Indels 22; Gaps 6;  
 QY 59 GCTTGGCCCCCGGCGCAAGAGCACTACTGCTGAAGCACTGTGTGGGCGAGCGCG 118  
 DB 136 GCGCTTGTGTCTCTCGACATCAAGAGCGCGAGTTCCGAGAGATGCTCAACTGA 195  
 QY 119 AGCTGAGGGCTTGCGCCCTGAACCCGGGCTGTGTGAAGACCGCGGAGGCTGCAAGCA 178  
 DB 196 CGGTGGGCAACGACGCTGCGCGCGGAGGTGTGTGAAGTGGCGGGAAACAAGCGCG 255  
 QY 179 TCATGAAGAGCTGCAAGCCCGCTTGCAGACCGGCAACGAGAGCTGCGCAGCTGTACA 238  
 DB 256 TCGTGAAGGTCTTTGAGGGCAGTGGGATGAGCCCAAGAACGCACTGCGAGTTCA 315  
 QY 239 ACACCGTGGCCACCTGTACTGTGTGCAAGCGGCACTGAGGTCCGCGCAACCAAGA-- 296  
 DB 316 CGGGGCACTATGCGCATCCCGTTCAGAGAGACATGCTGGGCGCGCTTCAACGCGCT 375  
 QY 297 -GGCCCTGGAACAAGATCGAGAGAGCAGAACAAAGTCCAGAGAAAGACCGCAGAGGCCA 355  
 DB 376 CGGGCGCGCCCTGCAAGAGGCGCGGCTCATGGCCAGAGCTTCTCTGCAATCCAG 435  
 QY 356 AGAGAGCGCAGCGCAAGGTGAGCCCAAGATCACTGCTGCAAGAACCTGCAAGGCGCAGA 415  
 DB 436 GCCAGCCCATCAACCCGTCAAGCGTGTATCCCCGAGAGATGATCCAGAGGGGATCA 495  
 QY 416 TGGTGACAGAGCGCATACGCCCGCCGACCTTGAAGCGCTGGGTGAAGGTATCGAGAGA 475  
 DB 496 GCGGCATCAACAGATGAATGATGCGCCGTGGCCAGAAAGTCCCGCTTCAACGCGCG 555  
 QY 476 AGGCTTCAAGCCCGAGGTGATCCCATGTTCACGCGCCTGAGCGAGGCGCCACCCGCC 535  
 DB 556 CCGGTGCGCGCAACAAGATTTGGCGGCAATCTGCAAGGAGGCGCGCTGTGGCGCC 615  
 QY 536 AGGACTGAACGATGTTTGAACCGTGGGCGGCCACCAAGGCGCCCATGCAATGCTGA 595

DB 616 TCCAGAGCAGCAAGAGGTGAGAGACCAACAGAGAGCAACTTCCGATCGTGTTCGGG 675  
 QY 596 AGAGACCATCAACAGAGAGGCGCGGATGGAGACCGCTGCAACCCCGTGCAGGCGGCG 655  
 DB 676 CCATGGGTGTCAACATGAGAGCTGCGAGTTCTTCCGCGAGACCTTCGAGGAAACGGCT 735  
 QY 656 CCGTGCCCGCCGCGCAGATGCGCGACCCCGCGGAGCAGATCGCGGCGCCACACAGA 715  
 DB 736 CGATGAGAGGTGTACAGCTGTTCTTCAACTGCGCAACCCACAGATGAGCGGATCA 795  
 QY 716 CCTTGAGAGAGATGCGCTGATGACCAAGCAACCCCGCTGCGCGGCGCATCT 775  
 DB 796 TCACGCGCGGTG---GCGCTCAGAGCGCGGATGACTCGCTGACAGTGCAGATGC 852  
 QY 776 ACAAGCGGATGATCTCTGAGGCT-----GAAACAATGTGGGATGTACAGCCCG 829  
 DB 853 ACGTGTGTATCTTACGAGACATGTGTGTGTGATGCGCGATGCTTGGGTGATCTCG 912  
 QY 830 TGAGCATCTGAGACATCCGCGCAGGCGCCCAAGAGCCCTTCCGAGCTACGTGACGCGCT 889  
 DB 913 CGGCGCGGAGAGGTGCTGGGCGTGTGTTACCCGCGTACATGTAACAGATTCG 972  
 QY 890 TCTTCAAGACCTGCGCGCGAGAGCGCAACCAAGAGATGAAAGTGAATGACCGAGA 949  
 DB 973 CGACCATCTACAGAGGTGCGGCGCGT---TCGAGGCGAGAAACGGCTGATCAACGACG 1029  
 QY 950 CCTGCTGTGAGACAGCGCAACCCGAGCTGCAAGACCATCTGCGGCTCTCGGCGCG 1009  
 DB 1030 TCCCATCTCTCAAGATGCGCAACAGACATCAACCACTCATCTACAGGCT 1089  
 QY 1010 GCGCACCTTGAAGAGATGATGACCGCTTCCAGAGGCGTGGGCGCGCCCGCGCAAGG 1069  
 DB 1090 ACATCAACGAGGCGCAGATTAACGTCAGACAGGCTGCAACAGGAGATTAACCGC 1149  
 QY 1070 C---CGGCTGTGGCGCAGCGCATGAGCCAGCGCAACAGGTGAATGATGACGA 1126  
 DB 1150 CCATCAAGCTGTGCTGCTGCTGAGGTGAGGAGGCGCATCGCGAGGCGATGA 1209  
 QY 1127 AGAGCACTTCAAGGCGCGCGCGCAACGTCAAGTCTTCAACTGCGCGCAAGAGGCGC 1186  
 DB 1210 CGCGGAGACACATCGGAGGTGTGAGACAGCTGTACCACTACCGCATTCGCGAAG 1269  
 QY 1187 ACATGCGCAAGACTGCGCGCGCGCGCGCAAGAGGCTGTGAAGTGGCGCAAGAGG 1246  
 DB 1270 ACAGATGCGCATGAAGGCTGTGTGTGTGAGGAGGCGCTGTCCCGAGGACAAAGCTCG 1329  
 QY 1247 GCCACAGATGAAGGACTGCAACGAGCGCGCAAGCTTCTGGGCAAGATCTGGGCCA 1306  
 DB 1330 CCTGAGGTTCTGAGAGAGTTCAAGAGAGATTCATCAACGAGGCTTCAACAGATGCG 1389  
 QY 1307 GCGCAAGAGGCGCGCGCGCAACTTCTGCAAGAACGCAAGGAGCGCGCGCGCGCGC 1366  
 DB 1390 GCGACATCTACAGATCGCTGACATCGCTGAGAGCTTGTG---CAAGTCCCAAG 1445  
 QY 1367 TGCCCAACGCGCGCGCGCGAGAGCTTGTGCTTCAAGAGACCAACCCCGCGCAAGC 1426  
 DB 1446 AGCTCTTCAAGCTATCAAGAGAAACCTTCAACAACTACTACAGAGGAGGCAAGN 1505  
 QY 1427 AGAGCCCAAGGA 1439  
 DB 1506 AGAGGACACAGTA 1518  
 RESULT 10  
 CK215205/c  
 LOCUS CK215205 746 bp mRNA linear EST 09-DEC-2003  
 DEFINITION FGAS027158 Triticum aestivum FGAS: Library 6 CAP GARE 1 Triticum  
 aestivum cDNA, mRNA sequence.  
 ACCESSION CK215205  
 VERSION CK215205.1 GI:39621309  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)

**ORGANISM** Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 746)  
**REFERENCE** Allard, P., Crosby, W.L., Danyluk, J., Budes, F., Frick, M., Gaudet, D.,  
Geneslwin, B., Graf, R., Gullick, P., Hrycan, L.D., Larchoe, A.,  
Linker, M.G., McCarthy, E.L., Morrey, A., Nuzak, I., Nilsson, D.,  
Pernikies, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)  
**TITLE** Contact: Mm L Crosby  
**JOURNAL** Bioinformatics  
**COMMENT** University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_estecs.usask.ca  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [13,662].  
Plate: LB6016 row: K column: 15.

**FEATURES**  
**SOURCE**  
Location/Qualifiers  
1..746  
/organism="Triticum aestivum"  
/mol\_type="rRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown  
(50%) and leaf (50%) tissues from wheat cultivar Norstar  
after short exposure times to low temperature in the light  
and in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20°C from wheat cultivar Norstar after  
short exposure times to low temperature in the light and  
in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20°C from wheat plants were transferred to 4°C  
for 7 days. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low  
temperature exposure. The last 6 populations: After 7 days  
of growth at 20°C, wheat plants were transferred to 4°C in  
the dark. 1cm crown sections and green leaf tissue were  
separately harvested after 1, 3, and 6 hours of low  
temperature exposure. First strand synthesis in this  
library was done in the presence of methylated dCTP  
thereby protecting from internal cleavage with NotI. In  
addition, this library used a primer for second strand  
synthesis that annealed to an artificial sequence (RNA  
oligo) added before first strand synthesis. Therefore when  
sequences from EST generated from this library will be  
masked for vector and adaptor sequences, an additional  
masking step will have to be included to mask this RNA  
oligo that is common to all clones (sequence  
CGACTGGACGACGAGCACTGCATGACTGAAGAGTAGAA)."

	Query Match	5.3%	Score 80.2;	DB 7;	Length 746;
Db	Best Local Similarity	49.7%;	Pred. No. 0.00018;		
	Matches 261;	Conservative 0;	Mismatches 258;	Indels 6;	Gaps 2;
Oy	191	TCGAGCCCCGCTCGAGACCGGACCAGAGAGCTCCGAGCTGTACAACCGTGCCA	250		
Db	670	TGCGCTCAAGCTCATCGGCGCAAGAAGCCATGAGATGCGCGCTTACACCGGTA	CTT		
Oy	251	CCCTGTACTGGGTGACCGCGGCATTCAGAGTCCGCGACCAAGAGGCGCTTGACA	310		
Db	610	CCCTCTCTGCCACCGCGCTCGCCATCCCAGAGACGACCACTTTGGCCATTGACAT	551		
Oy	311	TCGAGAGAGACAGAACATGTTCCAGACGAAGAACCCAGACAGCCAGAGGCGCAC	370		

Db	Accession	Length	Source
Db	550 ACCCGGAGAACTACGACTGGGGCTGCCGTGTCATTCGAGGAAGGCCGGGTGGCGGACAAAGA	491	
Qy	371 AGTGAGCCGAACTACCCCATCGTGCAGAACTGACAGGCGCAGATGATGACCAAGGCCA	430	
Db	490 TCGACTTCCCGAGAGGGCCCGGCGCTGGCTGGACGCGCTGTGAGAGACAGAGGCCA	431	
Qy	431 TCAGCCCCCGCACCTTAAAGGCTTGGGTGAAAGTGATTCGAGAGGAAGGCTTCAGCCCCG	490	
Db	430 ACACGGGACCTTTCGACTTGTGTCTTGTGAGACGCCGACAAAGAACACTTACTTCAATCAC	371	
Qy	491 AGGTGATTCGCCATGTTTCAACGCCCTTGAGCGAGAGGCGCCACCCCGCAGACTTGAACAGA	550	
Db	370 ACAGGCGCTTCATGAACTGTCTCAAGCTCCGCGGCTCTCTCGGTACAGAC---AACAACG	314	
Qy	551 TGTGAAACACCGTGGGCGGCCACCAAGCGCCCATTCGAGATGCTGAAGAGACACATCAACG	610	
Db	313 TCTGGAACGAGCTCGTGTGTCTCCCGCGCCGCCCATTCGCGAATGATATCGCTACT	254	
Qy	611 AGAAGCGCGCGAGTGGGACCGCTCTGCACCCCGTTCGAGCGCGCGCC---CGTGGCCCCG	667	
Db	253 ACCCGGACTTGTCTCTTCGACCTCAACAAAGGCGCTTCGCGCGCGACACAGCGGCTGAGATCT	194	
Qy	668 GCCAGATGCGGACCCCGCGCGAGCAGCAATTCGCGCGCGCACCA	712	
Db	193 GCCAGCTCCCGTGGCGGACGACATCACCTCTTGGCGCGCGGCCA	149	
RESULT 11			
AY106831			
LOCUS	AY106831	2299 bp	mRNA 1linear HTC 16-OCT-2002
DEFINITION	Zea mays PC0142079 mRNA sequence.		
ACCESSION	AY106831		
VERSION	AY106831.1		
KEYWORDS	GI:21209909		
SOURCE	HTC.		
ORGANISM	Zea mays		
	Zea mays		
	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.		
	1 (bases 1 to 2299)		
REFERENCE	Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.		
AUTHORS	Maize Mapping Project/DuPont Consensus Sequences For Design of Overgo Probes		
TITLE	Unpublished (2002)		
JOURNAL	2. (bases 1 to 2299)		
REFERENCE	Coe, E.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (25-APR-2002) Maize Mapping Project. University of Missouri, Columbia, MO 65211, USA		
JOURNAL	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.		
COMMENT	Location/Qualifiers		
	1..2299		
FEATURES	/organism="Zea mays"		
source	/mol_type="mRNA"		
	/db_xref="MaizEDB:638374"		
	/db_xref="taxon:4577"		
	/clone_idb="Maize Mapping Project/DuPont Consensus Library"		
	/note="this sequence is part of a project of EST assemblies resulting from the application of public access to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"		



Query Match 5.3%; Score 79.8; DB 3; Length 2299;  
 Best Local Similarity 43.9%; Pred. No. 0.0022;  
 Matches 495; Conservative 0; Mismatches 617; Indels 15; Gaps 3;

122 TGAAGGCTTCGCTGCTGAACCCGCTGCTGAGAACCGCGAGGCTGCAAGATCA 181  
 767 TCGAGGCGCGGCGCATATGAGAGCATCTGTGATGCGACCTCTTATATAGACGCCA 826  
 182 TGAAGAGCTGCAAGCCGCTTGCAGACCGGACCGAGAGCTGCGCAGCTGTACACA 241  
 827 AGAGAGTGAACGCCATGAGACCCGCTGTGAAGCGAAGCAGAGTACGCCCTCGCA 886  
 242 CCGTGGCCACCTGTACTGTGCTGACGCGCGCATTCAGATCCGCGACACCAAGAGGCC 301  
 887 CGTGCGCGAGTGGCTGCGCCCGCAAGTCAAGGTCATCCGCGCGCCACCAATTCATCG 946  
 302 TGAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361  
 947 AGGAGAGAGTCAACTGCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1006  
 362 CGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421  
 1007 TGAACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1066  
 422 ACCAGGCGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481  
 1067 CCGTGGCGAAGATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1126  
 482 TCAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541  
 1127 ACAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186  
 542 TGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601  
 1187 TCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1246  
 602 CCATCA-----ACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 655  
 1247 CCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1306  
 656 CCGTGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 715  
 1307 TCGTCTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1366  
 716 CCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 775  
 1367 ACATGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1426  
 776 ACAAGCGGTGATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 835  
 1427 CCGTCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1486  
 836 TCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 895  
 1487 ACCTTCACAGCGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1546  
 896 AGACCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 949  
 1547 TGTAGCGGATCACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606  
 950 CCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1009  
 1607 CCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1666  
 1010 GCGGCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1069  
 1667 TGTTCCTCAAGATCACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726  
 1070 CCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1129  
 1727 AGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1786

QY 1130 GCACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1189  
 DB 1787 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1843  
 QY 1190 TCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1236  
 DB 1844 CCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1890

## RESULT 12

CK211108/c

LOCUS

DEFINITION

FGAS022942 Triticum aestivum FGAS: library 6 CAP GATE 1 Triticum

aestivum cDNA, mRNA sequence.

ACCESSION

CK211108

KEYWORDS

SOURCE

ORGANISM

Triticum aestivum (bread wheat)

REFERENCE

AUTHORS

Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,

Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Larocque, A.,

Links, M.G., McCarthy, E.L., Monroy, A., Nazari, I., Nilsson, D.,

Penttilä, C., Roach, D.L. and Sarhan, F.

Functional Genomics of Abiotic Stress in Wheat and Canola Crops

Unpublished (2003)

Contact: Wm L Crosby

Bioinformatics

University of Saskatchewan, Department of Computer Science

1C101 Engineering Building, 57 Campus Drive, Saskatoon,

Saskatchewan, S7N 5A9, Canada

Tel: 306 966 1769

Fax: 306 966 2033

Email: fgas\_esteas.usask.ca

This sequence is the direct result of the Base calling software

Phred (default parameters). It is the raw base calls. To aid in the

identification of the high quality insert the software Lucy

(default parameters) has been run on this sequence. Lucy identified

the region [12,750].

Plate: 168001 row: A column: 20.

## FEATURES

source

1..1126

/organism="Triticum aestivum"

/mol\_type="mRNA"

/db\_xref="taxon:4565"

/clone\_lib="Triticum aestivum FGAS: library 6 CAP GATE 1"

/note="Organ: Crown and leaf; Vector: pCMV\_SPORT6; Crown

(50%) and leaf (50%) tissues from wheat cultivar Norstar

after short exposure times to low temperature in the light

and in the dark. 12 mRNA populations were combined before

constructing the library. The first 6 populations: After 7

days of growth at 20°C from wheat cultivar Norstar after

short exposure times to low temperature in the light and

in the dark. 12 mRNA populations were combined before

constructing the library. The first 6 populations: After 7

days of growth at 20°C, wheat plants were transferred to 4°C in

the dark. 1cm crown sections and green leaf tissue were

separately harvested after 1, 3, and 6 hours of low

temperature exposure. First strand synthesis in this

library was done in the presence of methylated dCTP

thereby protecting from internal cleavage with NotI. In

addition, this library used a primer for second strand

synthesis that annealed to an artificial sequence (RNA

oligo) added before first strand synthesis. Therefore when

sequences from EST generated from this library will be

masked for vector and adaptor sequences, an additional



## ORIGIN

masking step will have to be included to mask this RNA  
 oligo that is common to all clones (sequence  
 CGACTGAGCAGCAGGACACTGATGACTGAGAGTAGAAA)."

Query Match 5.2%; Score 78.4; DB 7; Length 1126;  
 Best Local Similarity 48.1%; Pred. No. 0.00038;

Matches 285; Conservative 0; Mismatches 301; Indels 6; Gaps 2;

124 GAGGGCTTGGCCCTGAAACCCGCGCTGCTGAGACCGCGGAGGCTGCAAGCAGATCAG 183  
 763 GAGATCACCGCCCAACCAACCATGAAACCTGATGACGACGCGGAGCGGCGCACTTC 704  
 184 AAGCAGCTCAGACCCCGCCCTGACAGACCGGACCGAGAGCTGGCGACGCTTCAACACC 243  
 703 CTCACATCTCTGTCAAGCTCATCGCGCCAGAGAACCATGAGATCGCGCTTACACC 644  
 244 GTGGCCACCTGTACTGCTGTCACGCGCGCATGAGTCCGCGACACCAAGAGGCGCTTC 303  
 643 GCGTACTCTCTCTCTCCGACCGCGCTCGCATCCCGACGACCGCATCTTGGCCATG 584  
 304 GACAGATGAGAGAGAGAGACAAAGTCCACAGACAGAACCCAGAGCGCAAGAGGCC 363  
 583 GACATCAACCGGAGAACTACGAGCTGGGGCTCCGTCATCGAAGAGCGCGCGCTGGCG 524  
 364 GACGGCAAGTGAAGCCAGAACTACCCCATGCTGACAGAACTGACAGGCGCAGATGTGCAC 423  
 523 CACAGATGACTTCTCGCGAGGCGCGCGCTGCGGTGAGACGCGCTGCGTAGAGAC 464  
 424 CAGGCGCATAGCGCCCGCAACCTGAAAGCGCTGGGTGAAGGTGATGAGAGAGAGCGCTTC 483  
 463 GAGGCCAACCAACGCGACCTTCCGACTTGTCTTGTGTGAGACGCCCAAGAGCAATCTCTC 404  
 484 AGCGCCGAGGTATCCCATGTTTACCGCCCTGAGAGAGGCGCGCAACCCCGACGACTTC 543  
 403 AACTACCAAGAGCGCTCATGAAAGCTGTCAACCTGCGCGCGCTCTCGCTGACGAC--- 347  
 544 AACAGATGTTGAACACCGTGGCGCGCCACAGCGCGCCCATGAGATGCTGAAGAGACACC 603  
 346 AACAGCTCTGAAACGCTCTCGGTGCTCCCGCGAGCGCCCATGAGAGGATCATCTC 287  
 604 ATCAACGAGAGAGCGCGCGAGTGGACCGGCTGACCGCGTGCAGGCGCGGCC---CGTG 660  
 286 CGCTCTTACCGGACCTTCTGCTCGACCTCAACAAGGCGCTCGCGCGCGACGCGCGTTC 227  
 661 GCCCGCGGCAAGTGGCGGACCGCGCGGAGAGACATCGCGCGCGCCACCA 712  
 226 GAGATGCGCAGGCTCCCGTGGCGGACGCGCATCACCTCTGCGCGCGCGCA 175

RESULT 13  
 CK163121 1064 bp mRNA linear EST 05-DEC-2003  
 LOCUS FGAS015739 Triticum aestivum FGAS: Library 4 Gate 8 Triticum  
 DEFINITION aestivum cDNA, mRNA sequence.  
 ACCESSION CK163121  
 VERSION CK163121.1 GI:38993028  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooideae; Triticeae; Triticum.  
 1 (bases 1 to 1064)

REFERENCE  
 AUTHORS Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
 Geneswein, B., Graf, R., Gulick, P., Hyman, L.D., Laroche, A.,  
 Links, M.G., McCarthy, E.L., Monroy, A., Musak, I., Nilsson, D.,  
 Penetier, C., Roach, J.L. and Sarhan, F.  
 Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
 Unpublished (2003)  
 COMMENT  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science

## FEATURES

source

1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033  
 Email: fgas.est@cs.usask.ca  
 This sequence is the direct result of the Base calling software  
 phred (default parameters). It is the raw base calls. To aid in the  
 identification of the high quality insert the software Lucy  
 (default parameters) has been run on this sequence. Lucy identified  
 the region [1,834].  
 Plate: L4B07 row: C column: 11.  
 Location/Qualifiers  
 1..1064

## ORIGIN

Query Match 5.1%; Score 77.4; DB 7; Length 1064;  
 Best Local Similarity 48.0%; Pred. No. 0.00058;  
 Matches 284; Conservative 0; Mismatches 302; Indels 6; Gaps 2;

124 GAGGGCTTGGCCCTGAAACCCGCGCTGCTGAGACCGCGGAGGCTGCAAGCAGATCAG 183  
 282 GAGATCACCGCCCAACCAACCATGAAACCTGATGACGACGCGGAGCGGCGCACTTC 341  
 184 AAGCAGCTCAGACCCCGCCCTGACAGACCGGACCGGAGAGCTGCGAGCTGTATCAACACC 243  
 342 CTCACATCTCTGTCAAGCTCATCGCGCCAGAGAGACATGAGATCGCGCTTACACC 401  
 244 GTGGCCACCTGTACTGCTGTCACCGCGCATGAGTCCGCGACACCAAGAGGCGCTTC 303  
 402 GCGTACTCTCTCTCTCGCGCACCGCGCTCGCATCCCGACGAGCGGACCATCTTGGCCATG 461  
 304 GACAGATGAGAGAGAGAGACAAAGTCCACGACAGAAACCCAGAGCGCCCAAGAGGCC 363  
 462 GACATCAACCGGAGAACTACGAGCTGGGCTGCGCTGTGATGAGAGGCGCGCGTGGG 521  
 364 GACGAGAGTGAAGCAGAACTACCCCATGCTGACAGAACTGACAGGCGCGAGATGTGCAC 423  
 522 CACAGATGACTTCTCGCGAGGCGCGCGCTGCGGTGAGACGCGCTGCGTAGAGAC 581  
 424 CAGGCGCATAGCGCCCGCAACCTGAAACGCTGGGTGAAGGTATGAGAGAGAGGCTTC 483  
 582 GAGGCCAACCAAGGACCTTTCAGCTTCTGTTGAGAGCGGACCAAGAGCAACTACCTTC 641  
 484 AGCGCCGAGGTATCCCATGTTTACCGCGCTGAGACGAGGCGGACCCCGCAGAGCTTC 543  
 642 AACTACCAAGAGCGGCTTCAAGAGCTGTCAAGCTTGGGCGGCTCTCGCTTACGAC--- 698  
 544 AACAGATGTTGAACACCGTGGCGCGCCACAGCGCGCCATGACAGTGTGAAGAGACACC 603  
 699 AACAGCTCTGAAAGGCTCTCTGCTGCTCCCGCGAGCGCCCATGAGAGATGATCATCTC 758  
 604 ATCAACGAGAGAGCGCGCGAGTGGACCGCGCTGACACCCGCTGACAGGCGCGGCC---CGTG 660  
 759 CGCTACTACCGGAGACTTGTCTGACCTCAACAAGGCGCGCGCGGACGACGCGCTTC 818

QY 661 GCCCCCCGCGCATGCGCGACACCCCGCGGACGACATCGCGCGCCACCA 712  
 Db 819 GAGATGTGCGAGGTCCCGCTCGCGCGACGCGATACCCCTCTGCGCGGCGCGCA 870

RESULT 14  
 CG351928 786 bp DNA linear GSS 26-AUG-2003  
 CG3AG53TH ZM 0.7-1.5 KB Zea mays genomic clone ZMMMA0754J09,  
 LOCUS genomic survey sequence.  
 DEFINITION  
 ACCESSION CG351928  
 VERSION CG351928.1 GI:34269194  
 KEYWORDS  
 SOURCE GSS.  
 ORGANISM Zea mays  
 Zea mays  
 Buxaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 786)  
 Whitefaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,  
 Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
 Clark,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other GSSs: CG3AG53TV  
 COMMENT  
 TITLE  
 JOURNAL  
 COMMENT  
 CONTACT: Cathy Whitefaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitefaw@tigr.org  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..786  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_1db="ZM-0.7-1.5\_KB"  
 /note="Vector: pBSK-; Site\_1: HincII, 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN  
 Query Match 5.1%; Score 77; DB 9; Length 786;  
 Best Local Similarity 46.1%; Pred. No. 0.0067;  
 Matches 334; Conservative 0; Mismatches 385; Indels 6; Gaps 2;

QY 579 CGCCATGACAGATGCTGAAGACACATCAACGAGGAGCGCGAGTGGAGCCGCTGCA 638  
 Db 776 CGCCACGCGCTCTGTGAGGACAGCTGAGCGCGACGTCGCGCAGCTGCAACAGATG 717  
 QY 639 CCCCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 698  
 Db 716 GGAGGAGGGCGAGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657  
 QY 699 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 758  
 Db 656 CGCGCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597  
 QY 759 GCCCGTGGCGCATCTACAGCGGTGATCTCTGCGCGCTGAAACAAGATGTCGCGAT 818  
 Db 596 GCGCGGTGAGCGCGCGCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537  
 QY 819 GATCAAGCCCGGTGAGCATCTGAGCATCCGCCAGAGGCGCCCAAGAGCCCTTCGCGACTA 878  
 Db 536 CAGCAGCTCTCCGATCCAGAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 477  
 QY 879 CGTGAAGCGCTTCTTCAAGACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 938  
 Db 476 GGAAGGAGGATGCGCGCGCGCTGCTGCGCTGCTGCGCGCGCGCGCGCGCGCG 420

QY 939 GATGACCGAGACCTCTGTGTCAGACGCAACCCCGCATCTGAAGACCATCTGCGCGC 998  
 Db 419 CCTGAGGATGAGGAGAGCTGTGTCTGTGCGGTCAACCGCG---GCTTGCGCTGCGCGGCGC 363

QY 999 TCTCGAGCCCGCGCGCGCACCTTGAAGAGATGATACCGCTTCCAGAGGCGTGGCGGCC 1058  
 Db 362 CGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 303

QY 1059 CGGCGCAAGAGCGCGCGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1118  
 Db 302 GTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 243

QY 1119 GATGAGAGAGCAACTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1178  
 Db 242 CGCGCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 183

QY 1179 GAGAGGCGCATGCGCGCAAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1238  
 Db 182 GCAGGAGGACATCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 123

QY 1239 CAAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1298  
 Db 122 CATTGTCGAGACCAAGTGGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 63

QY 1299 CTGGC 1303  
 Db 62 CAGGC 58

RESULT 15  
 CK163513 1138 bp mRNA linear EST 05-DEC-2003  
 LOCUS FGAS016142 Triticum aestivum FGAS: Library 4 Gate 8 Triticum  
 DEFINITION  
 ACCESSION CK163513  
 VERSION CK163513.1 GI:38933817  
 KEYWORDS  
 SOURCE Triticum aestivum  
 ORGANISM Triticum aestivum (bread wheat)  
 Buxaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 1138)  
 Allard,F., Crosby,W.L., Danyluk,J., Budes,F., Frick,M., Gaudet,D.,  
 Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,  
 Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilsson,D.,  
 Peniket,C., Roach,J.L. and Sarhan,F.  
 Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
 Unpublished (2003)  
 CONTACT: Wm L Crosby  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science  
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

FEATURES  
 source  
 Location/Qualifiers  
 1..1138  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4565"  
 /clone\_1db="Triticum aestivum FGAS: Library 4 Gate 8"  
 /note="Organ: Crown and leaf; Vector: pCMV.SPORT6;  
 Conditions for growth: Seeds were germinated in a  
 water-saturated mix (50% black earth and 50% ProMix) in a  
 growth chamber for 7 days under an irradiance of 200 mmol

m-2 sec-1. The temperature was maintained at 20 degrees C with a 15-hr photoperiod under a relative humidity of 70%. After this period watering of plants was stopped. Four time points were sampled during a two week period; the first after wilting was observed and the last, two weeks later, consisted of live crown and leaf tissue (leaf tissue that was yellow was not included in sampled material). First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

## ORIGIN

```
Query Match      5.1%; Score 76.8; DB 7; Length 1138;
Best Local Similarity 47.9%; Pred. No. 0.00074;
Matches 283; Conservative 0; Mismatches 303; Indels 5; Gaps 2;

QY 124 GAGGGCTTGCCTGAACCCCGCTGAGAGCCCGAGGGCTGCAAGCAGATCATG 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 232 GAGATCAGCGCCACACCAACCATGATGAGACCTGATGACGCTCGGCGGACGAGGCCAGTTC 291

QY 184 AAGCAGCTGCAAGCCCGCTGCAAGACCGGACCGAGAGCTGCGACGCTGTACAAACC 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 292 CTCACATGCTGCTCAAGCTCATCGGCGCCAGAGAACCATGAGATCGGCGCTTACACC 351

QY 244 GTGGCCACCTGTACTGCGGTGACGCGCGGATGAGGTCCGCGACCAAGAGAGGCCCTG 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 GGCTACTCTCTGTCTCCACCGCGCTGCGCATCTCCGACGACGGACCATCTTGGCCATG 411

QY 304 GACAAGATCGAGAGAGAGCAAGTCCACAGAGAACCCAGAGGCCAAGAGAGGCC 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 412 GACATCAACCGGAGAACTACAGAGCTGGGGCTGCCGTGATCGAAGAGGCCGCGTGGCG 471

QY 364 GACGGCAAGGTAGCCAGAACTACCCCATGTGACAGACTGCAAGGCCAAGATGTGAC 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 472 CACAAGATGACTTCCGCGAGGGGCCGGGCGTGCCTGAGACGCGCTGCTGAGAGAC 531

QY 424 CAGGCCATGAGCCCCCGCACCTGAAACGCTGAGTGAAGTGAATCGAGAGAGAGCCTTC 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 532 GAGGCCAACACAGGACCTTCACTTCTGTGTGACGCCGACAAAGACAACTACCTC 591

QY 484 ACCCCGAGGTATCCCATGTTCACCGCCCTGAGCGAGAGGCGCACCCCCAGACTG 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 592 AACTACCAAGACGCGCTCATGAAAGCTGTCAAGTGGGGGCTCTCTCGGCTACGAC--- 648

QY 544 AACACGATGTGAACAACCGTGGCGGACACAGGCCGCCATGCAATGCTGAAGACACC 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 649 AACACCCCTGTGAACGGCTCCGTCGTGCTCCCGCCGACGCCCCATGCGCAAGTACATC 708

QY 604 ATCAACGAGAGAGCCCGCAGAGTGGACCGCTGCAACCCGTGAGAGGCCGGC--CCGTGG 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 709 CGCTACTACCGGACTTCTCTCTCGAGCTCAACAGGCCCTCGCCGACACCAAGCGGTG 768

QY 662 CCCCCGCGAGATGCGGAGACCCCGGCGAGGAGCATGCGCGGCGCCACCA 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 769 AAGTCTGCCAGTCCCGCTCGGCGAGCGGATCACCTTGTGCGCGCGGCCCA 819
```

Search completed: March 13, 2005, 00:13:49  
Job time : 5318.82 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:04:55 ; Search time 52.7746 Seconds

(without alignments)  
5502.310 Million cell updates/sec

Title: US-09-475-704A-4

Perfect score: 2840  
Sequence: 1 atggcgccgcgcgcagcat.....ggcgccctgagccagtaa 1509

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO\_epool\_p/US09475704/runac\_10032005\_140222\_14839/app\_query.fasta\_1.3342  
-DB=PIR\_79 -QFMT=fasta -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0 -LIST=45  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -THR=45  
-LOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOM=ext -HEADSIZE=500 -MITLEN=0 -MAXLEN=200000000  
-USER=US09475704@cgn.1.160@runac\_10032005\_140222\_14839 -NCFU=6 -ICPU=3  
-NO WMAP -LANG=ENGLISH -NEG SCORES=0 -WAIT -DSPELACK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR\_79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARYS

Result No.	Score	Query Match	Length DB	ID	Description
1	2252.5	79.3	497	1	FOLJND gag polypotein -
2	2247	79.1	502	1	FOVWJ2 gag polypotein -
3	2237	78.8	512	1	FOVWH3 gag polypotein -
4	2233	78.6	500	1	FOVWH4 gag polypotein -
5	2232.5	78.6	501	2	S54377 gag polypotein -
6	2230	78.5	500	1	FOVWLV gag polypotein -
7	2227	78.4	500	1	A44001 gag polypotein -
8	2227	78.4	500	2	S33979 gag polypotein -
9	2204.5	77.6	505	2	T01677 gag polypotein -
10	2204.5	77.4	498	2	T09436 gag polypotein -
11	2179.5	76.7	506	1	A38068 gag polypotein -
12	2018.5	71.1	478	1	FOVWVL gag polypotein -
13	1963	69.1	508	1	FOLJSI gag polypotein -
14	1403.5	49.4	554	2	S46346 gag polypotein -

15	1396	49.2	510	1	FOLJTM gag polypotein -
16	1396	49.2	521	2	S08435 gag polypotein -
17	1390	48.9	521	1	FOLJCA gag polypotein -
18	1388	48.9	521	2	S53091 gag polypotein -
19	1385	48.8	521	1	FOLJST gag polypotein -
20	1381.5	48.6	507	2	S04237 gag polypotein -
21	1376	48.5	507	2	T11559 gag polypotein -
22	1374.5	48.4	522	1	FOLJG2 gag polypotein -
23	1372	48.3	521	2	S12152 gag polypotein -
24	1369.5	48.2	522	1	FOLJG6 gag polypotein -
25	1367.5	48.2	506	1	FOLJG3 gag polypotein -
26	1353.5	47.7	506	1	FOLJG5 gag polypotein -
27	1344	47.3	519	1	FOLJG4 gag polypotein -
28	1228.5	43.3	502	2	S28080 gag polypotein -
29	713.5	25.1	171	2	S52929 gag polypotein -
30	700	24.6	146	2	S60708 gag polypotein -
31	698	24.6	146	2	S60702 gag polypotein -
32	693	24.4	146	2	S60703 gag polypotein -
33	691	24.3	146	2	S60698 gag polypotein -
34	691	24.3	146	2	S60704 gag polypotein -
35	689	24.3	146	2	S60699 gag polypotein -
36	688	24.2	146	2	S60697 gag polypotein -
37	678	23.9	146	2	S60700 gag polypotein -
38	576.5	20.3	212	2	S03070 gag polypotein -
39	468	16.5	449	2	A45557 gag polypotein -
40	467	16.4	450	1	FOLJFP gag polypotein -
41	467	16.4	450	1	S23819 gag polypotein -
42	460	16.2	486	1	FOLJEV gag polypotein -
43	453	15.1	1106	2	S25162 gag polypotein -
44	429.5	15.1	1106	2	J00405 gag polypotein -
45	420	14.8	85	2	S49086 gag polypotein -

#### ALIGNMENTS

##### RESULT 1

FOLJND gag polypotein - human immunodeficiency virus type 1 (isolate NDK)

N:Alternate names: core protein p15; core protein p17; core protein p24

M:Contains: core protein p15; core protein p17; core protein p24

C:Species: human immunodeficiency virus type 1, HIV-1

A:Note: host Homo sapiens (man)

C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004

C:Accession: J00065

R:Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Gallibert, F.; Hampe, A.;

Gene 81, 275-284, 1989

A:Title: Nucleotide sequence of HIV-1-NDK: a highly cytopathic strain of the human immuno

A:Reference number: J00065; MWID:90034200; PMID:2806917

A:Accession: J00065

A:Molecule type: DNA

A:Residues: 1-497 <Spt>

A:Cross-references: UNIPROT:P18800; GB:M27323; NID:G328154; PIDN:AAA44868.1; PID:G328157

C:Genetics:

C:Superfamily: AIDS-related virus gag polypotein

C:Keywords: AIDS; core protein; immunodeficiency; polypotein

F:1-129/Product: core protein p17 #status predicted <C1>

F:130-389/Product: core protein p24 #status predicted <C2>

F:390-497/Product: core protein p15 #status predicted <C15>

##### Alignment Scores:

Pred. No.:	1,828-106	Length:	497
Score:	2252.50	Matches:	423
Percent Similarity:	91.49%	Conservative:	39
Best Local Similarity:	83.76%	Mismatches:	32
Query Match:	79.31%	Indels:	11
DB:	1	Gaps:	4

US-09-475-704A-4 (1-1509) x FOLJND (1-497)

QY	1	ATGGCGCCCGCGCGCAGATCTGCGCGAGAGACTGACCAAGTGGAGAGATCCGC	60
DB	1	MetGylalargAlaserValleuserGlyLysleuAapThrTrpGluArgIleay	20

```

QY 61 CTGGCCCGCGCGCAAGACACTACATCTGTAAGCACTGGTGGGCGGACCGCGAG 120
DB 21 LeuArgProGlyGlyIuYslyblyrYAlaLeuYshIleuIleTrrPaIasArGgln 40
QY 121 CTGAGAGGCTTCCGCGCTGAACCCCGGCTGTGAGAGCCGCGAGGGCTGCAAGCATC 180
DB 41 LeuGluArgPheThrIeuAenProGlyLeuLeuGluThrSerIuGlyCyblysgIinIle 60
QY 181 ATGAAGCAGCTGCAAGCCCGCGCTGCAAGCCGCAAGAGAGCTGGCGACCTGTACAAC 240
DB 61 IleGlyIleuLeuInProSerIleGlnThrGlySerGluIuIleArgSerLeuYrAan 80
QY 241 ACCGTGGCCCACTGTACTGCGTGCACGCGCGAGAGGTCCGCGCAACCAAGAGGCGC 300
DB 81 ThrValAlaThrIeuYrCybAlaHsGluArgIleGluValIlyAspThrIySgIuAla 100
QY 301 CTGAGCAAGATCGAGAGAGAGCAAAAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 101 ValGluIuYewerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120
QY 361 GCCGAGGCGAAGGTGAGGCGCAAGTACCCCATGTGTGCAAGCTTGCAAGGAGATGTG 420
DB 121 AspSerSerGlnValSerGlnAenYrProIleValGlnAenLeuGlnGlyIuIleVal 140
QY 421 CACGAGGCGATCAGCCCGCGCACTGTGAAGCGCTGGGTGAAGGTGATCGAGAGAGAGG 480
DB 141 HsGlnAlaIleSerProArgThrIeuAenAlaIrrPValIlyValIleGluIuYsAla 160
QY 481 TTCAGCGCGAGGTGATCCCATGTTCAACCGCGCTGAGGAGGCGCGCAACCCCGAGAG 540
DB 161 PheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAlaThrProGlnAen 180
QY 541 CTGAACACGATGTTGAACAACCGGCGCGCGCAAGCCCGCGCAAGTGTCTGAAGAGAC 600
DB 181 LeuAenThrMetLeuAenThrValGlyGlyHsGlnAlaIleMetGlnMetLeuYsGln 200
QY 601 ACCATCAACGAGGCGCGCGAGTGGAGCCGCTGCAACCCCGTGCAGGCGCGCGCGCG 660
DB 201 ThrIleAenAspGlnAlaIleGluTrrPaRPaIgleuHsIleProValHsAlaIleProVal 220
QY 661 GCCCGCGCGCAAGTGGCGCAACCCCGCGCGAGGAGATGCGCGCGCGCAACCGCTG 720
DB 221 AlaProGlyGlnMetArgIuProArgGlySerAspIleAlaGlyThrThrSerThrIeu 240
QY 721 CAGGAGCAGATCGCGTGGATGACCAAGCAACCCCGCGCGTGGCGCGCACTCAACAG 780
DB 241 GlnGluGlnIleAlaIrrPmetTrsSerAenProProIleProValGlyIuIleYrIyS 260
QY 781 CGGTGATCATCTGGGCGCTGAACAAGATCGTGGATGTACAGCCCGTGAAGATCTCTG 840
DB 261 ArgTrrPleIleIleuGlyLeuAenIySleValArgMetYrSerProValSerIleIeu 280
QY 841 GACATCCCGCAGGCGCGCGCAAGAGCCCTTCCGCGCACTACGTGAGACCGCTTCTCAAG 900
DB 281 AspIleArgGlnGlyProIySgIuProPheArgApyrYValAspArgPheYrIySthr 300
QY 901 CTGCGCGCGCAAGCGCGCAAGAGAGCGTGAAGAGCTGGATGAGAGAGAGAGAGAGAG 960
DB 301 LeuArgIaGluGlnIleAsenGlnAspYalIySAsnIrrPmetThrGluThrIeuAenVal 320
QY 961 CAGAAAGCCCAACCCGAGTGAAGACATCTGCGCGCTTCTGCGCGCGCGCGCAACCTG 1020
DB 321 GlnAenAlaAenProAspCyblySthrIleuYsAlaIleuGlyProGlnAlaThrIeu 340
QY 1021 GAGGAGATGATGACCGCGTGCAGAGGCGTGGGCGCGCGCGCGCAAGAGCGCGCTG 1080
DB 341 GlnGluMetMetThrAlaCybGlnIyAlaGlyGlyProGlyHsIySAlaIySValIeu 360
QY 1081 GCCGAGGCGATGAGCAG-----GCCAAGCGCTGAACATCATGATGACAGAGAGCAAC 1134
DB 361 AlaGluAlaMetSerGlnValThrGlySerAlaThrAlaValMetMetGlnIrrGlyAen 380

```

```

QY 1135 TTCAAGGCGCGCGCGCAACCTCAAGTCTTCAACTGCGCGCAAGAGGCGCACTCGCC 1194
DB 381 PheYsGlyProArgIySerIleIySleYsCybPheAenCybGlyYsGluGlyHsIthrAla 400
QY 1195 AAGAACTGCGCGCGCGCGCGCAAGAGGCGTGTGGAAGTGGCGCGCAAGAGGCGCAAG 1254
DB 401 LysAenCybArgAlaProArgIySleYsGlyStrIySleYsGlyArgIuGlyHsGln 420
QY 1255 ATGAAGCAGTGAACCGAGCGCGCAAGGCGCAACTTCTGGGCAAGATCTGGCGCAAG 1314
DB 421 MetIySAspCybThrGluArgGlnAlaAenPheLeuGlyIySleIrrPProSerHsIyS 440
QY 1315 GCGCGCGCGCGCAATTTCTGTGAAGACCGGAGAGAGAGAGAGAGAGAGAGAGAG 1374
DB 441 GlyArgProGlyAenPheLeuGlnSerAenProGlu-----ProThr 454
QY 1375 GCCCGCGCGCGCGAGAGCTTCCGCTTC---GAGAGACCAACCCCGCGCGCAAGAGAG 1431
DB 455 AlaProProAlaGluSerPheGlyPheGlyGluGluIleThrProSerGlnIySgln 474
QY 1432 CCAGAGAGCGCGAGCGCTTACCGGAGCGCGCTGACCGCGCTGCGAGCTGTTCGCGAGC 1491
DB 475 GlnYsAspIySgIuLeuYr-----ProIleuAlaSerIleuYsSerLeuPheGlyAen 492
QY 1492 GCGCGCGCGCGAGCAG 1506
DB 493 AspProSerSerGln 497

RESULT 2
FOVMA2
gag polyprotein - human immunodeficiency virus type 1 (isolate ARV-2)
N.Alternate names: core polyprotein
N.Contains: core protein p15; core protein p17; core protein p24
C.Species: human immunodeficiency virus type 1, HIV-1
A.Note: host Homo sapiens (man)
C.Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C.Accession: A03947
R.Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh
Science 227, 484-492, 1985
A.Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).
A.Reference number: A04003; MUID:85090453; PMID:2578227
A.Accession: A03947
A.Molecule type: DNA
A.Residues: 1-502 <SAS>
A.Cross-references: UNIPROT:P03349; GB:X02007; NID:G328658; PIND:AA59875.1; PID:G328661
C.Comment: Cleavage sites that yield the mature core proteins remain to be determined.
C.Genetics:
A.Gene: gag
C.Superfamily: AIDS-related virus gag polyprotein
C.Keywords: AIDS; core protein; immunodeficiency; polyprotein
F.1-134/Product: core protein p17 #status predicted <P17>
F.135-393/Product: core protein p24 #status predicted <P24>
F.394-502/Product: core protein p15 #status predicted <P15>

Alignment Scores:
Pred. No.: 3,44e-106 Length: 502
Score: 2247.00 Matches: 424
Percent Similarity: 90.78% Conservative: 39
Best Local Similarity: 83.14% Mismatches: 31
Query Match: 79.12% Indels: 16
DB: 1 Gaps: 5

US-09-475-704A-4 (1-1509) x FOVMA2 (1-502)
QY 1 ATGGGCGCGCGCGCGCAGACCTGCGCGCGGAGAGCTGGAACAAGTGGAGAAAGATCCGC 60
DB 1 MetGlyAlaArgAlaSerValIeuSerGlyGlyGluAenAspYrIySgIuYsIleArg 20
QY 61 CTGGCGCGCGCGCGCAAGAGCACTACATGCTGAAGACCTGGTGGCGCGAGCGCGAG 120
DB 21 LeuArgProGlyGlyIuYslyblyrYrYsLeuYshIleValIrrPaIasArGgln 40
QY 121 CTGAGAGGCTTCCGCGCTGAACCCCGGCTGTGAGAGCCGCGAGGGCTGCAAGCATC 180

```

[illegible]

```

Db      401  GYHLSIIEALIALysanCysArGAlAProARyglvslVcySTPrAcycseLIyRg    420
QY      1243  GAGGGCCACCAAGATTGAAGACTGCACCAGCGGCCAGGCCAACTTCTTGSGCAAGATCTGG    1302
Db      421  GIUGLylHSIGlMetLySaBPcYsthrGIuArgIlnIAAnPheLeugILyIsileTrp    440
QY      1303  CCCAGCACAAAGGGCGCGCCCCGGCAA CTTCGTGAGAACAACCGACGAGACCGCGCGCCC    1362
Db      441  ProSerTYrLYsgILyArGProlGLyAnPhneLueInserNzrGrProlV-----         456
QY      1363  ACCGTGCCCAACCGCGCCCGCCCCCGCCGAGAGCTTCGGCTTC-----GAGGAGCACACCCC    1416
Db      457  -----ProthrAlarProGrogluJuserPheargrHeaglGlugluTyThrThrPro    474
QY      1417  GCCCCCAAGCAGAGAGCCCAAGACCGCGAGCCCTTACCGCAGCCCTGACCGCCTGGCGC    1476
Db      475  SerGlnlySgIlngIuProlleApLySgIUleUtyr-----ProLeuthrSerDeuArxg    492
QY      1477  AGCCTGTTTCGCAAGCGCGCCCTTGAGCCAG    1506
Db      493  SerLeupheGLyAnAsAProSerSerGln    502

RESULT 3
FOVMH3
gag polyprotein - human immunodeficiency virus type 1 (isolate HTLV-III, BH10)
N:Alternate names: core polyprotein
N:Contains: core protein p15; core protein p17; core protein p24
C:Species: human immunodeficiency virus type 1, HIV-1
A>Note: host Homo sapiens (man)
C:Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
A:Accession: A03945
R:Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K.J.; Starcich, B.; Josephs, S.F.; Dorra-
nberger, J.A.; Papas, T.S.; Grayeb, J.; Chang, N.T.; Gallo, R.C.; Wong-Staal, F.
Nature 313, 277-284, 1985
A>Title: Complete nucleotide sequence of the AIDS virus, HTLV-III.
A:Reference number: A93353; MUID:85111123; PMID:2578615
A:Accession: A03945
A:Molecule type: DNA
A:Residues: 1-512 <RA>
A:Cross-references: UNIPROT:P03347; GB:M15654; GB:K02008; GB:K02009; GB:K02010; MID:g326
C:Comment: Cleavage sites that yield the mature core proteins remain to be determined.
C:Genetics:
A:Gene: gag
C:Superfamily: AIDS-related virus gag polyprotein
C:Keywords: AIDS; core protein; Immunodeficiency; polyprotein
F.1-132/Product: core protein p17 #status predicted <p17>
F.133-391/Product: core protein p24 #status predicted <p24>
F.392-512/Product: core protein p15 #status predicted <p15>

Alignment Scores:
Pred. No.:          Length:           512
Score:              Matches:            422
Percent Similarity:   89.69%             Conservative:   39
Best Local Similarity: 82.10%             Mismatches:     39
Query Match:        78.77%               Indels:         14
                          Gaps:           5

US-09-475-704A-4 (1-1509) x FOVMH3 (1-512)
QY      1 ATGGGCGCGCGCGCCAGCATCTCGCGCGCGAGAACTGGACAAGTAGTAATCCGC    60
Db      1 MetGlYAIArGalaserValIeusergdYLgIUleuAsPaRTpGlulysileArg    20
QY      61 CTGGCGCCCGGCGGGAAGAACACTACATGCTGAAGACCTGGGTGGGCGACGCGCGAG    120
Db      21 LeuAtgpProGLyGLylylsylsyTYrTLySLeuLysHISilevalITrpAlaserArglin    40
QY      121 CTGAGAGGCTTCGACCTGAACCCCGCGCTGCTGGAGACCGCGAGGCTGCAAGCAGATC    180
Db      41 LeuGlUArghPhealIValaenProoGLyLeuLeuGIUnthsereglULyCyArgrgIntle    60
QY      181 ATGAAGAAGCTGACGCGCGCGCTGCAGACCGGACCGAGAGAGCTGCGAGCTGTACAC    240

```

```

Db      61 LeuGIyGIInleuGIInProSerLeuGIInThrGIySerGIInIuLeuArgSerLeuTyraAsn 80
QY      241 ACCGTGGCCCACTGTACTGCGTGCAGCGCGGATCGGAGGTCCGCGACACCAAGAGGCG 300
      81 ThrValAlaThrLeuTyrcysValHisGlnArgIIeGIInIuLeuAspThrIySGIIuAla 100
QY      301 CTGGACAAGATCGAGAGGAGGAGCAACAAGTCCGACAGAAACCCAGAGGCGCAAGAG 360
      101 LeuAspIyIleGIInIuGIInIuGlnAsnIySerIyIyIyValGIInIuAlaIa 120
QY      361 GCCGAGCGC-----AAGTGAAGCCAGAACTACCCCTCGTGAAGAACTGCAAGGCG 411
      121 AspThrGIyHisSerSerGIInValSerGIInAsnTyrrProIleValGIInAsnIIeGIInIy 140
QY      412 CAGATGGTGCACCAAGGCGCATCAGCCCGCGACCTTGAAGCGCTGGTGAAGGTATCGAG 471
      141 GlnMetValHisGIInAlaIleSerProArgThrLeuAsnAlaTrpValIyValGIIn 160
QY      472 GAGAGAGGCTTCAGCGCCGAGGTATGCCATGTTGACCGCCCTGAGCGAGCGCGCGAC 531
      161 GlnIyAlaIlePheSerProGIInValIleProMetPheSerAlaLeuSerGIInIuGIInIyAlaThr 180
QY      532 CCCGAGACCTTGAACAAGATGTTGAACACCGTGGCGCGCACCGACCGCCATGACATG 591
      181 ProGIInAspLeuAsnThrMetLeuAsnThrValGIyGIyHisGIInAlaIleMetGIInMet 200
QY      592 CTGAGAGGACCATCTCAACGAGAGAGCGCGCGAGTGGGACCGCGCTGCAACCCCGTGCAGGCG 651
      201 LeuIySGIIuThrIleAsnGIInIuGIInIuAlaGIInIuTrpAspArgValHisProValHisAla 220
QY      652 GCGCCCGTGGCGCCCGGACGATGCGCGACCGCGCGCGCGCGAGCACTGCGCGCGCGCAC 711
      221 GIyProIleAlaIleProGIInIyIleMetArgIuProArgIySerAspIleAlaGIInIyThr 240
QY      712 AGCACCCTTGCAGAGAGAGATCGCTGGATGACAGCAACCGCCCGCGCGTGGCGAG 771
      241 SerThrIleuGIInIuGIInIleGIyTrpMetThrAsnAsnProProIleProValGIInIy 260
QY      772 ACTTCAAGCGGTGATCATCTCGTGGCGCTGAGCAAGATGCGTGGATGACAGCGCGCG 831
      261 ILeTyrcysArgIrrIleIleuGIInIyLeuAsnIyValIleArgIyIySerProThr 280
QY      832 AGCATCTGTGACATCGCGCGGCGCCCAAGAGACCGCTTCGCGACCTGAGCGCGCTTC 891
      281 SerIleLeuAspIleArgGIInIyProIySGIIuProPheArgAspTyrrValAspArgPhe 300
QY      892 TTCAAGACCTTCGCGCGCGAGAGCGCGACCGCAAGAGCTGAAGACTGATGACGAGACC 951
      301 TyrrIySerThrLeuArgAlaGIInIuAlaSerGIInIuValIyAsnTrpMetThrGIInThr 320
QY      952 CTGCTGTGTGCAAGAGCGCAACCGGACTGCAAGACCATCTGCGGCGCTCGCGCGCGCG 1011
      321 LeuLeuValGIInAsnAlaAsnProAspCysIySerThrIleLeuIyAlaLeuIyProAla 340
QY      1012 GCCACCTTGAAGAGATGATGACCGCTTGCAGCGCGCGTGGCGCGCGCGCGCGCAAGGCG 1071
      341 AlaThrIleuGIInIuMetMetThrAlaCysGIInIyValGIyGIyProGIyHisIyVal 360
QY      1072 CGCGGCTGGCGCGGAGCGCATGAGCGAG---GCCAACAGCGTGAACATCATGATGAGAG 1128
      361 ArgValIleuAlaGIInIuMetSerGIInValThrAsnThrAlaThrIleMetMetGIInArg 380
QY      1129 AGCAACTTCAAGAGGCGCGCGGCAAGCTCAAGTCTTCAACTGCGGCGCAAGAGGCGCAC 1188
      381 GlyAsnPheArgAsnGIInArgIyMetValIyCysPheAsnCysGIyIySGIIuGIyHis 400
QY      1189 ATCCGCAAGAACTGCGCGCGCGCGCGCAAGAGGCGCTCTGGAAGTGGCGGCAAGAGGCGC 1248
      401 ThrAlaAsnAsnCysArgAlaProArgIyIySGIIyCysTrpIyCysGIyIySGIIuGIy 420
QY      1249 CACGAGTGAAGAGTGCAGCGAGCGAGCGAGCGCAACTTCTGCGGCAAGATCTGGCGCGCGC 1308
      421 HisGIInMetIyAspCysThrGIInArgIInAlaAsnPheLeuGIyIySerIleTrpProSer 440

```

```

QY      1309 CACAAAGGCGCGCGCGCAACTTCTTGCAGAAACCGCAAGAGCGCGCGCGCGCACCGG 1368
      441 TyrrIySGIIyArgProGIyAsnPheLeuGIInSerArgProGIInProThrAlaProPhe 460
QY      1369 -----CCACCGCGCGCGCGCGCGAGAGCTTCGCG-----TTCGAG 1404
      461 LeuGIInSerArgProGIInProThrAlaProProGIInIuGIInSerPheArgSerGIyValGIIn 480
QY      1405 GAGACACCGCGCGCGCGCGCAAGAGAGCGCGCAAGAGCGCGCGCTTACCGGAGCGCGCTG 1464
      481 ThrThrIleProProGIInIySGIIuGIInIuProIleAspIySGIIuLeuTyrr-----ProIleu 498
QY      1465 ACCGCGCTTGCAGCGCTGTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGAG 1506
      499 ThrSerLeuArgSerLeuPheGIyAsnAspProSerSerGIIn 512

RESULT 4
FOVWHA
gag polyprotein - human immunodeficiency virus type 1 (isolate CDC-451)
N:Alternate names: core polyprotein
M:Contains: core protein p15; core protein p17; core protein p24
C:Species: human immunodeficiency virus type 1, HIV-1
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 09-Jul-2004
C:Accession: A25523
R:Deebl, S.M.; Kalyanaraman, V.S.; Casey, J.M.; Srinivasan, A.; Andersen, P.R.; Devare,
Proc. Natl. Acad. Sci. U.S.A. 83, 8380-8384, 1986
A>Title: Molecular cloning and primary nucleotide sequence analysis of a distinct human
A:Reference number: A94136; MUID:87041461; PMID:3490666
A:Accession: A25523
A:Molecule type: DNA
A:Residues: 1-500 <DES>
A:Cross-references: UNIPROT:P05887; GB:M1316; NID:G326459; PID:AAAA4306.1; PID:G326462
C:Genetics:
A:Gene: gag
C:Superfamily: AIDS-related virus gag polyprotein
C:Keywords: core protein; polyprotein
F:1.132/Product: core protein p17 #status predicted <P17>
F:1.133-391/Product: core protein p24 #status predicted <P24>
F:1392-500/Product: core protein p15 #status predicted <P15>

Alignment Scores:
Pred. No.: 1,75e-105 Length: 500
Score: 2233.00 Matches: 422
Percent Similarity: 90.35% Conservative: 37
Best Local Similarity: 83.07% Mismatches: 35
Query Match: 78.63% Indels: 14
DB: 1 Gaps: 5

US-09-475-704A-4 (1-1509) x FOVWHA (1-500)
QY      1 ATGGAGCGCGCGCGCGAGCATCTGCGCGGAGAGAGCTGAGCAAGTGGAGAAAGATCCGC 60
      1 MetGIyAlaArgAlaSerValIleuSerGIyGIyGIInIuAspArgTrpGIInIyIleArg 20
QY      61 CTGGAGCGCGCGCGCGCAAGAGCACTACATGTTAAGAACTCGTGGTGGCGAGCGCGGAG 120
      21 LeuArgProGIyGIyIySGIIuGIInIyArgIyIySerIleValIrrPalaserArgIy 40
      121 CTGAGAGGCTTGCCTGAAACCGCGCGCTGTGAGAGCGCGGAGGCTGCAAGCANATC 180
      41 LeuGIInArgPheAlaValAsnProGIyLeuLeuGIInIySerIySGIIyCysArgGIInIle 60
QY      181 ATGAAGAGCTGAGCGCGCGCGCGCGAGCAACCGCAAGAGAGTGGCGAGCTGTACAC 240
      61 LeuGIyGIInIuGIInProSerLeuGIInThrGIySerGIInIuIuLeuArgSerLeuTyraAsn 80
QY      241 ACCGTGGCCCACTGTACTGCGTGCAGCGCGGATCGGAGGTCCGCGACACCAAGAGGCGC 300
      81 ThrValAlaThrLeuTyrcysValHisGlnArgIIeGIInIuLeuAspThrIySGIIuAla 100
QY      301 CTGACAAGATCGAGAGGAGAGCAACAAGTCCAGAGCAAGAGCCAGAGCGCGCAAGAG 360

```



```

Db      101 LeuAspLysrIleGIuGIuGIuGlnAaenLysSerLysLysValaGlnGlnAlaAla 120
QY      361 GCCGAGGC-----AAGTGAGCCGAGAACTACCCCATCGTGCAGAACTGCGAGGC 411
Db      121 AspThGlyAsnSerSerGlnValSerGlnAaenLysProIleValGlnAaenLysGln 140
QY      412 CAGATGGTGCACGAGGCATCAGCCCGGCAACCCCTGAAGCGCTGGGGAAGATGATGAG 471
Db      141 GlnMetValHisGlnAlaIleSerProAaenLysLeuAaenLysValaIleGln 160
QY      472 GAGAAAGCCTTCAGCCCGGAGTGAATCCCATGTTACCGCTTGAAGCGAGGCGGCACC 511
Db      161 GIuLysAlaIlePheSerProGIuValIlePheMetPheAlaIleLeuSerGIuValaThr 180
QY      532 CCCCAGGAGCTTGAACGATGTTGAACCCGTGGGCGGCAACGAGCCGCGCATGCAATG 591
Db      181 ProGlnAspLeuAaenLysMetLeuAaenLysValaGlyHisGlnAlaIleMetGlnMet 200
QY      592 CTGAAGAGACCACTCAACGAGGAGGCGGAGTGGAGCCGCTGACCCCGTGCAGGCGC 651
Db      201 LeuLysGlnThrIleAaenGIuGlnAlaIleGlnTTPAspArgLeuHisProValHisAla 220
QY      652 GGCCTCGTGGCCCGGCGGAGATGCGCGAGCCCGGCGGAGCGACATCGCGCGCGCAC 711
Db      221 GlyProIleAlaProGIuLysMetArgGluProArgGlySerAspIleAlaGlyThrThr 240
QY      712 AGCACCCTTCAGAGAGATCGCTGATGATGACCAACCCCGCTGCCCTGGGCGAC 771
Db      241 SerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnAsnProProThrProValGlyGlu 260
QY      772 ATCTCAAGAGGCGGATCATCTCGGCGCTGAACAAATGTTGGTGAAGTGAAGCCCGG 831
Db      261 IleTrpLysArgTrpIleIleLeuGlyLeuAaenLysIleValaArgMetLysSerProIle 280
QY      832 AGCATCTCTGAGATCCGCGAGGCGCCAGAGAGCCCTTCGCGCATCACTGAGCCGCTTC 891
Db      281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspLysTrpValaAspArgPhe 300
QY      892 TTCAGAACCTTCGCGCGCGGAGCGGCGGAGCGGAGCTGAAGAACTGATGACCGAGCC 951
Db      301 TrpLysThrLeuArgHisGlnGlnAlaSerGlnGlnValLysAaenTrpMetThrGluThr 320
QY      952 CTGCTGGTGCAGAGCGCAACCCCGGCTGAAGACCATCTGCGGCTTCGCGCGCGCGC 1011
Db      321 LeuLeuValaGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340
QY      1012 GCCACCTTGAGAGATGATGACCGGCTGCTGAGCGGTGGCGCGCGCGGCAAGGCC 1071
Db      341 AlaThrLeuGIuGlnMetMetThrAlaCysGlnGlyValaGlyGIuProGIuHisLysAla 360
QY      1072 CGCGTGTGGCGGAGCGGATGAGCCAG--GCCAACAAGCTGAACATCATGATGACAGAG 1128
Db      361 ArgValaIleAlaGlnAlaIleMetSerGlnValaThrAsnSerAlaThrIleMetMetGlnArg 380
QY      1129 AGCAATTCAGAGGCGCGCGGCGGCAACGTCAAGTGTTCATCTGCGGCAAGAGCGCCAC 1188
Db      381 GlyAaenPheArgArgGlnGlyLysThrValaLysCysPheAsnLysGlyLysGlnGlyHis 400
QY      1189 ATGCGCAAGAACTGCGCGCGCGCGCGGCAAGAGGCGTGTGAGAGTGGCGGCAAGAGGCGC 1248
Db      401 IleAlaArgAsnCysLysValaIleProArgLysLysGlyCysTrpLysCysGlyArgGlnGly 420
QY      1249 CACCAAGATGAGATGCAACCGAGCGGCAAGCGCCCACTTCCTGGGCAAGATCTGGCGCAC 1308
Db      421 HisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440
QY      1309 CACCAAGGCGCGCGCGGCAACTTCCTGCAAGAACCGGAGAGCGCGCGCGCGCCACCGG 1368
Db      441 HisLysGlyArgProGIuLysAaenPheLeuGlnSerArgProGln----- 454
QY      1369 CCCACCGCGCGCGCGCGGAGGCTTCGCGCTTC-----GAGGAGACCAACCCCGCGCGCC 1422
Db      455 ProThrAlaProProGIuLysSerPheArgPheGlyAspGlnThrThrThrProSerGln 474

```

```

QY      1423 AACGAGAGCCCAAGAGACCGCGAGCCCTACCGCGAGCCCTTGAACCGCGCTTGGCAGCCG 1482
Db      475 LysGlnGluProArgAspLysGlnLys-----ProLeuAlaSerLeuArgSerLeu 492
QY      1483 TTCGCGAGCGCGCGCGCTGAGCCAG 1506
Db      493 PheGlyAaenAspProSerSerGln 500

RESULT 5
S54377
gag polyprotein - human immunodeficiency virus type 1
C.Species: human immunodeficiency virus type 1, HIV-1
C.Date: 15-Jul-1995 #sequence revision 01-Sep-1995 #text_change 03-Jul-2004
C.Accession: S54377
R.Theodore, T.; Buckler-White, A.J.
submitted to the EMBL Data Library, July 1989
A.Reference number: S54377
A.Accession: S54377
A.Status: preliminary
A.Molecule type: genomic RNA
A.Residues: 1-501 <TH>
A.Cross-references: UNIPROT:P12495; EMBL:M2639; NID:G329377; PIDN:AAA45365.1; PID:G3293
C.Superfamily: AIDS-related virus gag polyprotein
C.Keywords: polyprotein

Alignment Scores:
Pired. No.: 1,85e-105 Length: 501
Score: 2232.50 Matches: 424
Percent Similarity: 90.37% Conservative: 36
Best Local Similarity: 83.30% Mismatches: 34
Query Match: 78.61% Indels: 15
DB: 2 Gaps: 5

US-09-475-704A-4 (1-1509) x S54377 (1-501)
QY      1 ATGGGCGCGCGCGGAGATCTGCGGCGGAGAACTGGAACAAGTGGAGAAATCCGCG 60
Db      1 MetGlyAlaArgAlaSerValaIleuSerGlyLysLeuAspAlaTrpGluLysIleArg 20
QY      61 CTGGCGCGCGCGGAGAGACATCATGTGTAAGCACTGTGTGGCGGCGGAG 120
Db      21 LeuArgProGIuGlyLysLysLysTrpArgLeuLysHisLeuValaTrpAlaSerArgGlu 40
QY      121 CTGAGAGCGCTTCGCTGAAACCCGCGCTGTGTGAGACCGCGGAGCGCTGCAAGATG 180
Db      41 LeuGluArgPheAlaLeuAaenProGIuLeuGlnThrSerAspGlyCysLysGlnIle 60
QY      181 ATGAAGCGCTGCAGACCGCGCTGCAAGCCGCGGAGAGCTGCGAGCTGTGAAC 240
Db      61 IleGlyLysLeuGlnProAlaIleArgTrpLysSerGlnGluLeuArgSerLeuPheAsn 80
QY      241 ACCGTGACCACTGTATGTCGTGCAACCGCGGATGAGTCCGAGACCAAGAGAGCGC 300
Db      81 ThrValaIleThrLeuLysCysValaHisGlnArgIleGlnValaLysAspThrLysGlnAla 100
QY      301 CTGCAACAAGTGCAGAGAGAGCAAGAACTGTCAGCAAGAGACCAAGAGAGAGAG 360
Db      101 LeuGluLysMetGlnGlnGlnGlnAaenLysSerLysAaenLysValaGlnGlnAlaIle 120
QY      361 GCCGAG-----GGCAAGTGAAGCAACCTAACCTATGTCGACG 408
Db      121 AlaAspAlaGlyAaenAsnSerGlnValaSerGlnAaenLysProIleValaGlnAaenLysGln 140
QY      409 GCGCAGATGTGCAACGAGCATAGCCCGCGCACTGGAACGCTGGGTGAAGTGAATC 468
Db      141 GlyLysMetValHisGlnAlaIleSerProArgTrpThrLeuAaenAlaTrpValaLysValaIle 160
QY      469 GAGGAGAAAGCCTTCAGCCCGAGGATGATCCCATGTTTCAACCGCGCTTGAAGAGAGCGCGC 528
Db      161 GIuGluLysAlaIlePheSerProGIuValIlePheMetPheSerAlaLeuSerGIuLysAla 180
QY      529 ACCCGCGAGACCTGAACAACATGTTGAACACCGTGGGCGGCGACCAAGCGCGCATGCGAG 588

```



Db 161 GluYbAlaPheSerProgluVal11leProMetPheSerAlaLeuSerGluGlyAlaThr 180  
 QY 532 CCCCGAGACTTGAACACGATGTTGAACCCGTGGGGGCGCCACGAGCCCGCATGCGAGG 591  
 Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMet 200  
 QY 592 CTGAAGACACCATCAACGAGAGGCGCGAGTGGAGACCGCTGACCCCGCGAGGCG 651  
 Db 201 LeuYbGluThrIleAsnGluGluAlaIleGluThrAspThrValHisProValHisAla 220  
 QY 652 GGGCGCGTGGCGCGCGCGAGATGCGAGCCCGCGCGGACGACATCGCGCGCGCAC 711  
 Db 221 GlyProIleAlaProGluGlnMetArgGluProArgGlySerAspIleAlaGlyThrThr 240  
 QY 712 AGCACCCTTCAGAGACGATCGCTGATGACGACACCCCGCGCGCGCGCGCGCAC 771  
 Db 241 SerThrLeuGlnGlnGlnIleGlyTrpMetThrAsnAsnProProIleProValGlyGlu 260  
 QY 772 ATCTCAAGCGGGGATCATCTCGGGGCTGAACCAAGATGTCGATGTCAGAGCCCGG 831  
 Db 261 IleTrpYbAspTrpIleIleLeuGlyLeuAsnLysIleValArgMetCysSerProThr 280  
 QY 832 AGCATCTTGAGACATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 891  
 Db 281 SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTrpValAspArgPhe 300  
 QY 892 TTCAGACCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951  
 Db 301 TyrLysThrLeuArgIleGluGlnIleAsnGlnGluValLysAsnTrpMetThrGluThr 320  
 QY 952 CTGCGTGTGTCAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1011  
 Db 321 LeuLeuValGlnAsnLysAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAla 340  
 QY 1012 GCCACCTTGAGAGAGATGATGACCGCGCTGCGAGCGCGCGCGCGCGCGCGCGCG 1071  
 Db 341 AlaThrLeuGluGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAla 360  
 QY 1072 CGCGTGTGTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1128  
 Db 361 ArgValIleuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArg 380  
 QY 1129 AGCACTTCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1188  
 Db 381 GlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnCysGlyLysGlnLysHis 400  
 QY 1189 ATCGCAAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1248  
 Db 401 IleAlaArgAsnCysArgIleProArgLysGlySerTrpLysCysGlyLysGlnGly 420  
 QY 1249 CACCGATGAGAGATGTCACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1308  
 Db 421 HisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSer 440  
 QY 1309 CACAAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1368  
 Db 441 TyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGln----- 454  
 QY 1369 CCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1422  
 Db 455 ProThrAlaProProGluGlnIleSerPheArgSerGlyValGluThrThrTrpProSerGln 474  
 QY 1423 AAGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1482  
 Db 475 LysGlnGlnLysProIleAspLysGlnLeuTrp-----ProLeuThrSerLeuArgSerLeu 492  
 QY 1483 TTTCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1506  
 Db 493 PheGlyAsnAspProSerSerGln 500

RESULT 7  
 A44001  
 gag polyprotein - human immunodeficiency virus type 1 (strain YU-2)

N:Alternate names: core polyprotein  
 N:Contains: core protein p15; core protein p17; core protein p24  
 C:Species: human immunodeficiency virus type 1, HIV-1  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
 C:Accession: A44001  
 R:Li, Y.; Hu, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.  
 J. Virol. 65, 6587-6600, 1992  
 A:Title: Complete nucleotide sequence, genome organization, and biological properties of  
 A:Reference number: A44001; M01D:93021387; PMID:1404605  
 A:Accession: A44001  
 A:Molecule type: DNA  
 A:Residues: 1-500 <Lit>  
 A:Cross-references: UNIPROT:P35962; GB:M93258  
 C:Genetics:  
 A:Gene: gag  
 C:Superfamily: AIDS-related virus gag polyprotein  
 C:Keywords: AIDS; core protein; immunodeficiency; polyprotein  
 F:1-132/Product: core protein p17 #status predicted <P17>  
 F:133-391/Product: core protein p24 #status predicted <P24>  
 F:392-500/Product: core protein p15 #status predicted <P15>

Alignment Scores:  
 Pred. No.: 3,51e-105 Length: 500  
 Score: 2227.00 Matches: 423  
 Percent Similarity: 89.96% Conservative: 34  
 Best Local Similarity: 83.27% Mismatches: 37  
 Query Match: 78.42% Indels: 14  
 DB: 1 Gaps: 5

US-09-475-704a-4 (1-1509) x A44001 (1-500)

QY 1 ATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
 Db 1 MetGlyAlaArgAlaSerValLeuSerAlaGlyIleuAspLysTrpGluLysIleArg 20  
 QY 61 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
 Db 21 LeuArgProGlyGlyLysArgGlnTrpArgLeuLysHisIleValTrpAlaSerArgGln 40  
 QY 121 CTGGAGGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
 Db 41 LeuGlnArgPheAlaValAspProGlyLeuLeuGlnThrTrpSerGlnGlyCysArgGlnIle 60  
 QY 181 ATGAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
 Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGluLeuArgSerLeuTrpAsn 80  
 QY 241 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
 Db 81 ThrValAlaThrLeuTrpCysValHisGlnLysIleGluValLysAspThrLysGluAla 100  
 QY 301 CTGACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 Db 101 LeuGluLysIleGluGlnGlnGlnAsnLysSerLysLysValGlnGlnAlaIleAla 120  
 QY 361 GCCACGCGC-----AAGGTGAGCGAGAACTACCCCATGTGTCAGAGGCGC 411  
 Db 121 AspThrGlyAsnSerSerGlnValSerGlnAsnTrpProIleValGlnAsnLeuGlnGly 140  
 QY 412 CAGATGTCGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 471  
 Db 141 GlnMetValHisGlnAlaIleSerTrpArgThrLeuAsnAlaTrpValLysValGlu 160  
 QY 472 GAGAGGCGCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 531  
 Db 161 GluYbAlaPheSerProgluVal11leProMetPheSerAlaLeuSerGluGlyAlaThr 180  
 QY 532 CCCCGAGACTTGAACACGATGTTGAACCCGTGGGGGCGCCACGAGCCCGCATGCGAGG 591  
 Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMet 200  
 QY 592 CTGAAGACACCATCAACGAGAGGCGCGAGTGGAGACCGCTGACCCCGCGAGGCG 651

Db 201 Leuylsdlutlhrllleahnglulgluallalaglultrpharprg-leuhsiprovalhla 220  
QY 652 GGGCCCTGGCCCCGGCCGAGATCGCGACCCCGCGGAGGAGATCGCGCGCGCAC 711  
Db 221 GlyProilealProglgimelArggluProhrglYserhpielaelaglythrtr 240  
QY 712 AGACCCCTGAGAGAGAGATCGCTGTGATGACAGAACCCCCCGCTGGCGCGGAC 771  
Db 241 SerThrleuglnlunlunlleglytrpmetThrAsnAsnProProilleProvalGlyglu 260  
QY 772 ATCTACAGGGGGTGGATCATCTGGGGCTTGACAGATCGTGGATGATACAGCCCGG 831  
Db 261 llerlylsargrtrpillelleuglyleuamlylelvalargmetryserprothr 280  
QY 832 AGCATCTGAGACATCGCGCGCGGAGCCGAGAGACCTTCGCGAGCTGACCGCTTC 891  
Db 281 SerilleuamspillearglnglyProlysgluPropharprgAspYrValAsparphe 300  
QY 892 TTCAAGACCTTGGCGCGCGAGAGGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 951  
Db 301 TyrlyThrleuarglaleuglnlaserlngluvallyasanthrmetThrnglu 320  
QY 952 CTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
Db 321 LeuileuValGlnAsnAlaAsnProAspCylysthrilleuulysalaleuglyProala 340  
QY 1012 GCCACCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071  
Db 341 AlathrleuglnlunlunlmetethrAlaCysglnlyValGlyProlyhAslysa 360  
QY 1072 CGCGTCTGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128  
Db 361 ArgvalleuAlaglnlalemetserglnvalThrAsnserAlathrillemetethrArg 380  
QY 1129 AGCAATTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188  
Db 381 GlyAsnPharprgAsnGlnlunlunlunlunlunlunlunlunlunlunlunlunl 400  
QY 1189 ATGCGCAAGATCGCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1248  
Db 401 llealalyasnthrarglaleProahglnlyblysglyCysrplyCysglnlyleugly 420  
QY 1249 CACCGAGTGAAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308  
Db 421 HisglnmetlyAspCysThrngluarglnlaleAsnPheluglylylelletpProser 440  
QY 1309 CACAAGGGGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368  
Db 441 HislysglyArgProglYAsnPheluglnserArgProglu----- 454  
QY 1369 CCCACCGCGCGCGCGCGAGAGCTTCGCGCTTC-----GAGAGAGAGAGAGAGAGAG 1422  
Db 455 ProthrAlaProsergluglnserValarghphelgyluglnlunlunlunlunlunlunl 474  
QY 1423 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482  
Db 475 lysglnlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunl 492  
QY 1483 TTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506  
Db 493 PheglYserAspProserSerleu 500

## RESULT 8

S33979

gag polyprotein - human immunodeficiency virus type 1

C:Species: human immunodeficiency virus type 1, HIV-1

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004

C:Accession: S33979

R:Carlini, F.

submitted to the EMBL Data Library, November 1991

A:Reference number: S33979

A:Accession: S33979

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-500 <CAA>  
A:Cross-references: UNIPROT:O78240; EMBL:Z11530; NID:g60192; PIDN:CAA7621.1; PID:g60193  
C:Superfamily: AIDS-related virus gag polyprotein

## Alignment Scores:

Pred. No.:	3,51e-105	Length:	500
Score:	2227.00	Matches:	420
Percent Similarity:	90.16%	Conservative:	38
Best Local Similarity:	82.68%	Mismatches:	36
Query Match:	78.42%	Indels:	14
DB:	2	Gaps:	5

US-09-475-704a-4 (1-1509) x S33979 (1-500)

QY 1 ATGGGCGCGCGCGCGAGATCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlylyLeuAspArgGlylyValArg 20  
QY 61 CTGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
Db 21 LeuArgProglYglnlylylylylylylylylylylylylylylylylylylylylyly 40  
QY 121 CTGAGAGGCTTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
Db 41 LeuGlnArgPheAlaValAsnProglYleuLeuGlnThrSerGlnlyCysArgGlnlle 60  
QY 181 ATGAAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 61 LeuGlnleuGlnleuGlnleuGlnleuGlnleuGlnleuGlnleuGlnleuGlnleuGln 80  
QY 241 ACCGTGGCACCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
Db 81 ThrValAlaThrLeuYrCyValHisGlnArglleGlnlyleuAspThrlysgluAla 100  
QY 301 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspYlleglnlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunl 120  
QY 361 GCCACGCGC-----AAGGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 411  
Db 121 AspThrGlyHisSerSerGlnValSerGlnAsnThrProilleValGlnAsnilleGlnly 140  
QY 412 CAGATGGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471  
Db 141 GlnmetValHisGlnAlaLeuSerProArgThrLeuAsnAlaThrVallyValGln 160  
QY 472 GAGAGAGCTTCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531  
Db 161 GlnlyshAlaPheSerProglYValillePrometPheSerAlaLeuSerGlnlyAlaThr 180  
QY 532 CCCAGAGAGCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 591  
Db 181 ProGlnAspLeuAsnThrmetlleuAsnThrValGlyGlyHisGlnAlaAlaMetGlnmet 200  
QY 592 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651  
Db 201 Leuylsdlutlhrllleahnglulgluallalaglultrpharprg-leuhsiprovalhla 220  
QY 652 GGGCCCTGGCCCCGGCCGAGATCGCGACCCCGCGGAGGAGATCGCGCGCGCAC 711  
Db 221 GlyProilealProglglnlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunl 240  
QY 712 AGACCCCTGAGAGAGAGATCGCTGTGATGACAGAACCCCCCGCTGGCGCGGAC 771  
Db 241 SerThrleuglnlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunlunl 260  
QY 772 ATCTACAGGGGGTGGATCATCTGGGGCTTGACAGATCGTGGATGATACAGCCCGG 831  
Db 261 llerlylsargrtrpillelleuglyleuamlylelvalargmetryserprothr 280  
QY 832 AGCATCTGAGACATCGCGCGCGGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891

DB 281 SerIleLeuAspIleArgInglYProLysGluProPheArgAspTyrValAspArgPhe 300

QY 892 TTCAGAGACCCCTCGCGCCGAGAGGCGACCCGAGACGTAAGAACTGGATGACGAGCC 951

DB 301 TyrLysPThrLeuArgIleArgInglYAsnIleSerArgInglValLysAsnTrpMetThrGluThr 320

QY 952 CTGCTGTGTGCAGAAAGCCCAACCCCGACTGCAGAACATCTCTGCGCGCTCTCGGCCCGCGC 1011

DB 321 LeuLeuValGlnAlaAsnIleAsnProAspCysLysPThrIleLeuLysAlaLeuGluYProGlu 340

QY 1012 GCCACCCCTGAGAGAGATGATGACCGCTCTCCAGGGCGTGTGGCGGCCCGCCGACCAAGGCC 1071

DB 341 AATTrLeuGluGlnMetMetThrIleAcysGlnGlnValGlyGlyProGluYHISLysVala 360

QY 1072 CGCGGTGTGCGCCGAGGCGCATGAGCCAG---GCCAACAGCGTGAACATCATGATGACAGAG 1128

DB 361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnProAlaAsnIleMetMetGlnArg 380

QY 1129 AGCAACTTCMAAGGGCCCGCGCGCAACGTCAAGTCTTCACTGCGGCAAGAGGGCCAC 1188

DB 381 GlyLysPheArgAsnGlnGlnLysThrValLysCysPheAsnCyGlyLysGlnGluYHIS 400

QY 1189 ATCGCCAAAGACTGCGCGCGCGCCCGCCGCAAGAAAGGCGTGTGGAATGCGGCAAGAGGGC 1248

DB 401 IleaIleAspAsnCyAspArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGluY 420

QY 1249 CACCAAGATGAAGAGATGTGACCGAGGCGGCGCAACTCTCTGGGCGAAGATCGGCGCCAGC 1308

DB 421 HisGlnMetLysAspCysThrGlnArgInglAlaAsnPheLeuGlyLysIleTrpProSer 440

QY 1309 CACAAAGGCGCGCGCGCGCAACTCTCTGCAAGAACCGCACGAGCGCGCGCCCGCCACCGTG 1368

DB 441 HisLysGlyArgProGluYAsnPheLeuGlnSerArgProGlu----- 454

QY 1369 CCCACCGCGCGCGCGCGCGAGAGCTTCCGCTTC-----GAGAGACCAACCCCGCGCCCC 1422

DB 455 ProThrAlaProProGluGlnLysPheArgPheGlyGluGlnThrThrProSerGln 474

QY 1423 AAGCAGGAGGCCCAAGAACCGGAGCCCTAACCGCGAGCCCGCTGACCGCCCTCGCAGCGCTG 1482

DB 475 LysGlnGluProIleAspLysGlnMetCys-----ProLeuAlaSerLeuArgSerLeu 492

QY 1483 TTCGGCAGCGCGCGCGCGCTGAGCCAG 1506

DB 493 PheGlyAsnAspProSerSerGln 500

RESULT 9

T01667

gag polyprotein - human immunodeficiency virus type 1

C:Species: human immunodeficiency virus type 1, HIV-1

C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 09-Jul-2004

C:Accession: T01667

R:Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.

A:Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two iso

A:Reference number: Z14389; MUID:8645056; PMID:2424612

A:Accession: T01667

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-505 <Alt>

A:Cross-references: UNIPROT:P04594; EMBL:X03456; NID:g60228; PIDN:CAA28011.1; PID:g60222

C:Superfamily: AIDS-related virus gag polyprotein

Alignment Scores:

Pred. No.: 4.78e-104 Length: 505

Score: 2204.50 Matches: 422

Percent Similarity: 88.33% Conservativity: 32

Best Local Similarity: 82.10% Mismatches: 39

Query Match: 77.62% Indels: 21

DB: 2 Gaps: 6

US-09-475-704A-4 (1-1509) x T01667 (1-505)

QY	1	ATGGGCGCCGCGCCAGCATCTCGCGCGCGCGAAGCCTTGACAACAATGGGAGAAATATCCC	60
Db	1	MetGlyAlaArgAlaSerValLeuSerGlyGlyLysLeuAspAlaTrpGluLysIleArg	20
QY	61	CTGCGCCCGCGCGCGCAGAGCACTACATGTCTGAAGCACTGTATGTGGCCAGCCCGAG	120
Db	21	LeuArgProGlyGlyLysLysIleSerTrpTrpGluLysIleValTrpAlaSerArgGlu	40
QY	121	CTGAGAGGCGTTGCGCTTGAACTCCGCGCTCTGAGACCGCGGAGGCTGCAAGCATTC	180
Db	41	LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluTrpGlyGluGlySerGlnIle	60
QY	181	ATGAAGAGCTGCGACCGCGCTCGAAGCCCGGACCGGAGAGATGGGCGCGCTGTCAAC	240
Db	61	MetGluGlnLeuGlnSerThrLeuLysThrGlySerGluGluIleLysSerLeuTrpAsn	80
QY	241	ACCGTGGCCACCTCTACTCTCGGTGCAGCCCGGCATCGAGTCCGCGACACCAAGAGGC	300
Db	81	ThrValAlaTrpLeuTrpCysValHisGlnArgIleAspValIleAspThrLysGluAla	100
QY	301	CTGCAACAATTCGAGAGAGAGCAACAATGCCCAAGACCAAGAACCCAGAGGCCAAGAG	360
Db	101	LeuAspLysIleGluGluIleGlnAsnLysSerArgGlnLysThrGlnIleAlaAla	120
QY	361	GCC-----GACGCGAAGTGGAGCCAGACCACTACCCCATC	393
Db	121	AlaGlnGlnAlaAlaAlaAlaThrLysAsnSerSerValSerGlnAsnTrpProIle	140
QY	394	GTGCAGAACTGCGAGGCGCCAGATGTGCACCAAGCCATCAGCCCCCGACCTGAAGCC	453
Db	141	ValGlnAsnAlaGlnGlyIleMetIleHisGlnAlaIleSerProArgThrLeuAsnAla	160
QY	454	TGGGGAAGGTGATGAGAGAGAGGCGTTACGCGCCGAGGTATCCCATGTTCACCGCG	513
Db	161	TrpValLysValIleGluGluLysAlaPheSerProGluValIleProMetPheSerAla	180
QY	514	CTGAGCGAGGCGCGCCACCCCGAGCACTGAACAGATGTTGAACACCGTGGCGCGCAC	573
Db	181	LeuSerGluGlyAlaThrProGlnAspLeuAsnMetC MetLeuAsnIleValGlyGlnHis	200
QY	574	CAGCGCCGCGATGCGAGATGTCAGACAACATCAACGAGAGAGCGCGCGAGTGGAGACCC	633
Db	201	GlnAlaAlaMetGlnMetLeuLysAspThrIleAsnGluGluAlaAlaAspTrpAspArg	220
QY	634	CTGCAACCCGCGAGCGCGCGCCGCTGCGCCCGCGCGACATGCGCGACCCCGCGAGAC	693
Db	221	ValHisProValHisAlaGlyProIleProProGlyGlnMetArgGluProArgIleSer	240
QY	694	GACATGCGCGCGCCACCAAGCACTCTGACAGACAGATGCGCTGATGACACGAAACCC	753
Db	241	AspIleAlaGlyThrThrSerThrLeuGlnGlnGlnIleGlyTrpMetThrSerAsnPro	260
QY	754	CCCGGCGCGGCGGCGATCTACAGAGGCGTGAATCATCTGGGCGCTGAACAAGATCCG	813
Db	261	ProIleProValAlaGlyAspIleTrpLysArgTrpIleIleGlnGlyLeuAsnLysIleVal	280
QY	814	CGAGTGTACAGCCCGGTAGCATCTTGACATCCGCGAGGCGCCCAAGAGGCCCTTCCGC	873
Db	281	ArgMetTrpSerProValSerIleLeuAspIleArgGlnGlyProLysGluTrpPheArg	300
QY	874	GACTTACGTGAGCCGCTTCTTCAAGACCTTGCGCGCGCGAGAGCCCAACCAAGACGTGAAG	933
Db	301	AspTrpValAlaPargPhePheLysThrLeuArgAlaGluGlnAlaThrGlnIleValLys	320
QY	934	AACGTGATGACCGAGACCTCTGTGTGACAGACGCAACCCCGACTGCAAGACATCTCTG	993
Db	321	AsnTrpMetThrGluTrpThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeu	340
QY	994	CGCGCTCTCGGCGCCGCGCGCAACCTTGAGAGATGATGACCGCTTCCAGAGGCGTGGGC	1053
Db	341	LysAlaLeuGlyProGlyAlaThrLeuGluGluMetMetThrAlaCysGlnIleValGly	360

[illegible]

## RESULT 10

gag polyprotein - human immunodeficiency virus type 1 (strain JREFL)  
C:Species: human immunodeficiency virus type 1, HIV-1  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T09436  
R: Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,  
submitted to the EMBL Data Library, July 1996  
A: Reference number: Z16673  
A: Accession: T09436  
A: Status: preliminary; translated from GB/EMBL/DBJ  
A: Molecule type: DNA  
A: Residues: 1-498 <PAM>  
A: Cross-references: UNIPROT:O75754; EMBL:U63632; NID:G1465777; PID:G1465778  
C: Genetics:  
A: Gene: gag  
C: Superfamily: AIDS-related virus gag polyprotein

Alignment Scores:  
Pred. No.: 1.02e-103 Length: 498  
Score: 2198.00 Matches: 418  
Percent Similarity: 90.32% Conservative: 39  
Best Local Similarity: 82.61% Mismatches: 37  
Query Match: 77.39% Indels: 12  
DB: Gaps: 6

US-09-475-704A-4 (1-1509) x T09436 (1-498)

[illegible][illegible]

Db 421 MetlysaApCyThnGluArgGlnAlaenPheluGlyLysIleTrpProSerTyrls 440

Qy 1315 GGGCCGCCCGGCACTTCTCGAGAACCGCAGCGCCCGCCCGCCCGCCGCTGCCACC 1374

Db 441 GtAAGpProGlyLysenPheluGlnSerArgProGlu-----Prothr 454

Qy 1375 GCGCCCGCCCGGAGAGCTTCCGCTTC---GAGGAGACC---ACCGCCGCCCCAAGCAG 1428

Db 455 ALaPProGluLysSerPhenArgPheluGlyGluTrnAlaThrProSerGlnLysGln 474

Qy 1429 GAGCCCAAGAACCGGAGCGCCCTACCGCAGCGCCCGCCGCGCCGCGAGCTGTGGC 1488

Db 475 GtUProlleApLysGluMetTyLr-----ProleuTrnSerLeuArgSerleuPhely 492

Qy 1489 AGCGGCGCCCTGAGCCAG 1506

Db 493 AenAepProSerSerGln 498

RESULT 11

A38068

gag polyprotein - human immunodeficiency virus type 1 (strain MN)

N:Alternate names: core polyprotein

N:Contains: core protein p1, core protein p17, core protein p2; core protein p24; core p

C:Species: human immunodeficiency virus type 1, HIV-1

A>Note: host Homo sapiens (man)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004

C:Accession: A38068

R:Henderson, L.E.; Bowers, M.A.; Sowder II, R.C.; Serabyn, S.A.; Johnson, D.G.; Bess Jr

J:ViroL.66, 1856-1865, 1992

A:Title: Gag proteins of the highly replicative MN strain of human immunodeficiency viru

A:Reference number: A38068; MUID:9219415; PMID:1548743

A:Accession: A38068

A:Molecule type: protein

A:Residues: 1-506 <HEN>

A:Cross-references: UNIPROT:P05888

C:Genetics:

A:Gene: gag

C:Superfamily: AIDS-related virus gag polyprotein

C:Keywords: AIDS; core protein; immunodeficiency; polyprotein

F:1-134/Product: core protein p17 #status experimental <P17>

F:135-365/Product: core protein p24 #status experimental <P24>

F:366-379/Product: core protein p2 #status experimental <P2>

F:380-434/Product: core protein p7 #status experimental <P7>

F:435-450/Product: core protein p1 #status experimental <P1>

F:451-506/Product: core protein p6 #status experimental <P6>

Alignment Scores:

Pred. No.:	8.71e-103	Length:	506
Score:	2179.50	Matches:	413
Percent Similarity:	88.72%	Conservative:	43
Best Local Similarity:	80.35%	Mismatches:	37
Query Match:	76.74%	Indels:	21
DB:	1	Gaps:	6

US-09-475-704A-4 (1-1509) x A38068 (1-506)

Qy 4 GGGCGCGCGCAGAGATCTCTGCGCGGAGAGCTGGAAGTGGGAGAAATCCGCTG 63

Db 1 GtYAlaAhgAlaSerValLeuSerGlyGlyGluLeuAphArgTrpGluLysIleArgLeu 20

Qy 64 CGCCCGCGCGGAGAACACTACTAGCTGTGAAGCACTGATGTGGCGGAGCGGAGCTG 123

Db 21 ArgProGlyLysIleLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLysLeu 40

Qy 124 GAGGAGCTTGGCCCTGAAACCCCGGCTGCTGAGAACCGCCGAGGGCTGCAACAATCATG 183

Db 41 GtUAArgPhenAlaIleAenProGlyLeuLeuGlnTrnSerGlyGlyCybArgGlnIleLeu 60

Qy 184 AAGCACTGCAAGCGCGCTGCAAGCGGAGCGGAGAGCTGCGAGCGCTGTACAACACC 243

Db 61 GtYlneGlnPProSerLeuGlnThrGlySerGlnGluArgLysSerLeuTyIenThr 80

Qy 244 GTGGCAACCTGTACTGCGTGACGCGGCAATCGAGTCCGCGAACCAAGAGGCCCTG 303

Db	81	ValAlaThrLeuTyrCysValAlaHISGlnLysIleLysIleLysAspThrLysGlnLysLeu	100
Oy	304	GACAAGATCGAGAGAGAGACAGAAACAGTCCGACAGAAAGACCCAGACGACCAAGAGGCC	363
Db	101	GlnLysIleGlnGlnGlnGlnGlnAsnLysSerLysLysLysAlaGlnGlnAlaAlaLys	120
Oy	364	GACGGC-----AAGTGAAGCCAGAACTAACCCCATCTGCACGAACCTG	405
Db	121	ThrGlyAsnArgGlyAsnSerSerGlnValSerGlnAsnTyrProIleValGlnAsnIle	140
Oy	406	CAGGGCCAGATGATGTCACAGCCCATCAGCCCCCGGACCCCTGAAGCGCTGGGTGAAGGTG	465
Db	141	GlnGlyGlnMetValHISGlnAlaIleSerProArgTrpThrLeuAsnAlaTrpValLysAl	160
Oy	466	ATCGAGGAGAAAGGCGCTTCAGCCCGGAGGTATATCCCATGTTCAACGGCCCTAGCGAGGCG	525
Db	161	ValGlnGlnLysValAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGlnGly	180
Oy	526	GCCACCCCCCAGACCTGAACACGATGTTGAACACGTTGGCGGCGCACAGCGCCCATG	585
Db	181	AlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHISGlnAlaAlaMet	200
Oy	586	CAGATGCTGAAGACACCATCAACAGAGAGGCCCGCCGATGGAGACCGCTTCAACCCCTG	645
Db	201	GlnMetLeuLysGlnTrpIleAsnGlnGlnAlaIleGlnTrpAspArgLeuHISProAla	220
Oy	646	CAGGCGGGCGCCGAGGCCCGCGGACAGATGGCGCACCCCGCGGACAGCATCGCCGCG	705
Db	221	HISAlaGlyProIleAlaProGlyGlnMetArgLysProHISGlySerAspIleAlaGly	240
Oy	706	GCCACACGACCTCTGCAGAGACAGATGCGCTGAATGACACGAACCCCGCTGCCGTG	765
Db	241	ThrThrSerThrLeuGlnGlnGlnIleGlyTyrPheThrAsnAsnProProIleProVal	260
Oy	766	GCGCAGCATCTTCAGCGGTGATCATCTCGGGCTTGAAACAGATCTGGCGGATGTACAGC	825
Db	261	GlyGlnIleTyrLysArgTrpIleIleLeuGlnLysLeuAsnLysIleValAlaArgMetCysSer	280
Oy	826	CCCGTGAAGCATCTTCGACATCGCGCAGAGGCCCAAGAGAGCCCTTCGCGCATACGTGAG	885
Db	281	ProSerSerIleLeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAsp	300
Oy	886	CGCTTCTTCAAGACCTCTGCGCGCGGACAGGCGCACCGACGTAAGAACTGATGAC	945
Db	301	ArgPheTyrLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTrpMetThr	320
Oy	946	GAGACCTCTGCTGTGCAGAGCGCCACCCCGACCTGCAGAACCATCTGCGGCTCTCGGC	1005
Db	321	GlnThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGly	340
Oy	1006	CCCGGCGCCACCTTCGAGAGAGATATACCGCTGCCAGAGGGCGTGGCGCGCGCGCAC	1065
Db	341	ProAlaAlaThrLeuGlnGlnMetMetThrAlaCysGlnGlyValGlyIleProGlyHis	360
Oy	1066	AAGGCGCGGTCTGGCGGAGCGCATGACGAG--GCCAACACGCTGAACATCATGATG	1122
Db	361	LysAlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaAlaThrIleMetSer	380
Oy	1123	CAGAAGAGCACTTCAAGAGGCCCGCGGCGACGTCATCAAGTCTTCAACTGCGGCAAGAG	1182
Db	381	GlnArgGlyAsnPheArgAsnGlnAlaArgLysIleIleLysCysPheAsnCyGlyLysGln	400
Oy	1183	GGGCACATTCGCAAGAACTGCGCGGCGCCCGGCAAGAAAGGCTGTGGAAATGGCGGCAG	1244
Db	401	GlyHISIleAlaLysAsnCyAspAlaProArgLysArgGlyCysTrpLysCyGlyLys	420
Oy	1243	GAGGCGCACAGATGAAGACTGACACGAGCGGCGGCGCACTTCTTGGCGAAGATCTGG	1302
Db	421	GlnGlyLysGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrp	440
Oy	1303	CCGACGCCAAGAGGCGCGCCCGGCACTTCTGTGAGAAACGCGACGAGCCCGCGCGCCC	1362



Dh 441 ProSerCysAlaGlyArgProGlyAenPheProGlnSerArgThrGlu----- 456  
Qy 1363 ACCGTGCCCAACCGCCCGCCCGGAGAGCTTCGCTTC-----GAGGAGACCAACCC 1416  
Dh 457 -----ProThrAlaProProGlnGluSerPheArgPheGlyGluThrThrPro 474  
Qy 1417 GCCCGGAGAGAGGCCCAAG-----GACCGGAGAGCCCTACCGGAGCCCTG 1464  
Dh 475 TyrGlnGlySerGlnGlyLysGlnGluThrTrlLeuAspLysAspLeuThr-----ProLeu 492  
Qy 1465 ACCGCGCTGCGAGCGCTGTTCGCGAGCGCGCCCTGAGCCAG 1506  
Dh 493 AlaSerLeuLysSerLeuPheGlnAenAspProLeuSerGln 506

## RESULT 12

## FOVWVL

gag polyprotein - human immunodeficiency virus type 1 (isolate LV)

N:Alternate names: assemblin; core polyprotein; gag precursor

N:Contains: capsid antigen core protein p24CA; core protein p1; core protein p6; matrix

C:Species: human immunodeficiency virus type 1, HIV-1

C:Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004

C:Accession: A03948

R:Accession: M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.

Nature 313, 450-458, 1985

A:Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi

A:Reference number: A93355, NCID:8511157, PMID:2982104

A:Accession: A03948

A:Molecule type: DNA

A:Residues: 1-478 <MW>

A:Cross-references: UNIPROT:P03350

C:Gene: gag

C:Superfamily: AIDS-related virus gag polyprotein

C:Keywords: AIDS; blocked amino end; core protein; immunodeficiency; lipoprotein; myrist

F:2-478/Product: gag precursor (assemblin) #status predicted <GAG>

F:2-132/Product: matrix antigen core protein p17MA #status predicted <P17>

F:20-32/Region: nuclear location signal

F:110-114/Region: nuclear location signal

F:133-363/Product: capsid antigen core protein p24CA #status predicted <P24>

F:364-377/Product: core protein p2 #status predicted <CP2>

F:378-433/Product: nucleocapsid core protein p7NC #status predicted <CP7>

F:392-405/Region: zinc finger CCHC motif

F:413-426/Region: zinc finger CCHC motif

F:2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted

F:392,395,400,405/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

F:413,416,421,426/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

## Alignment Scores:

Prod. No.:	1.15e-94	Length:	478
Score:	2018.50	Matches:	382
Percent Similarity:	87.86%	Conservative:	45
Best Local Similarity:	78.60%	Mismatches:	40
Query Match:	71.07%	Indels:	19
DB:	1	Gaps:	5

US-09-475-704A-4 (1-1509) x FOVWVL (1-478)

Qy 1 ATGGGCGCCCGGCGAGCATCTGCGCGCGGAGAGCTGCAAGTGGAGAAATCCGC 60  
Dh 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTyrGlnLysLeArg 20  
Qy 61 CTGGCGCCCGGCGGAGCAAGCATCTGCTGAGAGACCTGTGTGGCGGAGCCGAG 120  
Dh 21 LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
Qy 121 CTGGAGGCGCTTGCCTTAACCCCGGCTGTGAGAGACCGCGAGGCTGCAAGATC 180  
Dh 41 LeuGlnArgPheAlaValAlaAspProGlyLeuLeuGlnThrSerGlyGlyCysArgGlnLe 60  
Qy 181 ATGAAGCAGTGCAGCCCGCTGAGAGACCGGAGAGCTGGCGAGCTGTGACAC 240  
Dh 61 LeuGlnLeuGlnProSerLeuGlnThrGlySerGlnGlnLeuArgSerLeuThrAsn 80

Qy 241 ACCGTGCCCAACCGCCCGCCCGGAGAGCTTCGCTTC-----GAGGAGACCAACCC 300  
Dh 81 ThrValAlaThrLeuThrCysValAlaGlnArgGlnGlnLeuLysAspThrLysLys 100  
Qy 301 CTGGACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Dh 101 LeuAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 120  
Qy 361 GCCGAGCGC-----AAGGTGAGCCAGCACTACCCCATCGTGCAGACCTGCAGAGGC 411  
Dh 121 AspThrGlyLysSerSerGlnValSerGlnValSerGlnValSerGlnValSerGln 140  
Qy 412 CAGATGTGCACAGAGCCATACAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471  
Dh 141 GlnMetValAlaGlnAlaAlaSerProArgThrLeuAsnAlaThrValValValGlu 160  
Qy 472 GAGAGAGCGCTTGCAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531  
Dh 161 GlnLysAlaLysSerProGlnValAlaLeuProMetPheSerAlaLeuSerGlnGlyAlaThr 180  
Qy 532 CCCGAGAGCTGAACAGATGTGACACCGTGGGCGGCGAGAGAGAGAGAGAGAGAGAGAG 591  
Dh 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlnGlnAlaAlaMetGlnMet 200  
Qy 592 CTGAAGACACCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651  
Dh 201 LeuLysGlnThrLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220  
Qy 652 GAGCCCGGCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 711  
Dh 221 GlyProLeuAlaProGlyGlnMetCysArgGlnProArgLysSerAlaAlaGlyThr 240  
Qy 712 AGCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771  
Dh 241 SerThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
Qy 772 ATCTCAAGCGGTGATCATCTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 831  
Dh 261 IleTyrLysArgTyrPheLysLysLysLysLysLysLysLysLysLysLysLysLys 280  
Qy 832 AGCATCTGAGCATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
Dh 281 SerLeuLeuAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 300  
Qy 892 TTCAAGACCTGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951  
Dh 301 TyrLysThrLeuArgAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 320  
Qy 952 CTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
Dh 321 LeuLeuValGlnAsnAlaAspProAspCysLysThrLysLysLysLysLysLysLys 340  
Qy 1012 GCCACCTGAGAGAGATGATGACCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071  
Dh 341 AlaThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 360  
Qy 1072 CGGTGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128  
Dh 361 ArgValLeuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrLysMetGlnArg 380  
Qy 1129 AGCAACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188  
Dh 381 GlnAspPheArgAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 400  
Qy 1189 ATGGCCAGAACTGCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1248  
Dh 401 ThrAlaArgAsnCysArgAlaProArgLysGlyCysThrPheCysGlyLysGlnGly 420  
Qy 1249 CACCAAGTGAAGAGCTGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305  
Dh 421 HisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuLysLysLysLysLys 440  
Qy 1306 AGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1365





## RESULT 14

gag polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)

C/Species: simian immunodeficiency virus SIVagm

A; Variety: isolate SAB-  
C: Date: 25-Dec-1994 #

```
C:\Date: 25-Dec-1994 #sequence_revision 14-Feb-1997 #text_change 26-Aug-1999
C:\Accession: S46346
```

R;Jin, M.J.; Hui, H

EMBO J. 13, 2935-2947, 1994

A;Title: Mosaic genome stru

A/Reference number: S46335; MUID:94298785; PMID:8026477

A/Accession: 546346

A:Molecule type: DN

A:Residues: 1-554 <JIN>  
A:molecule type: DNA

A;Cross-references: EME

A; Experimental source:

A; Note: the nucleotide sequence was submitted to the EMBL Data Library, De

C:Genetics:

A/Gene: gag

C:Superita  
C:Keyworo

### Keywords

Alignment Scores:

Pred. No.:

**Score:**

Percent Similarity:	65.18%	Conservative:	75
---------------------	--------	---------------	----

```
Best Local Similarity: 51.798  Mismatches: 122
Query Match: 48.428  Total: 73
```

Query Match:	49.42%	Indels:	73
DB:	2	Gaps:	14

DB:	2	Gaps:	14
-----	---	-------	----

US-09-475-704A-4 (1-1509) X S46346 (1-554)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

US-09-475-704A-4 (1-1509) X S46346 (1-554)

QY 1 ATGGGCCCCCGCCAGCATCTCTGCAGCGCAGAAAGCTGGACAAGTGGGAAGAATCCGC 60  
| : : : : :  
Db 1 MetClAlaseraanservalLeuSerGIAlArGLyLeuSPrAlpheIGLuserValArg 20

QY 61 CTGCGCCCCGGCGCAAGAACACTAATCTGAAGCACTGTGTGTGGCCAGCCGAG 120  
| : : : : :  
Db 21 LeuArGrProAnGIyILySLyLyTyTLyLeuArGrHISLeuValTPRAlasetLyGLu 40

QY 121 CTGAGAGGGCTTCGCCCTTGAAACCCCGGCGCTGTGGAAACCGCCGAGGCTGCAGACATC 180  
| : : : : :  
Db 41 LeuAsPrArGrPheSerLeuSerAlaAsnLeuIleuNIThrLySGluValValysILE 60

QY 181 ATGAGAGAGCTGCAGCCCGCCCTGCAGACCGGCAACCGAGAGAGCTGGCCACCCTGTACAAC 240  
| : : : : :  
Db 61 LeuSerValLeuLeuProLeuValProThrGLySerGIuAsnLeuIleAlaLeuPheAsn 80

QY 241 ACCGTGGCCAACCTGTACTGTGCTGCACGCGCGACTCGAGTGCAGTCCGCGCACCAAGAGAGCC 300  
| : : : : :  
Db 81 LeuCysCyValLeuAlaCyVILeIALeALeGLuLeLyValLyAsPrThrGIuGLuAla 100

QY 301 CTGGACAAGATCCAGAGAGAG-----CAGAAC 327  
| : : : : :  
Db 101 LyseAlaLySeValLySGluGIuValProAlaGLuMetThrGluserAlaThrAlatInrSer 120

QY 328 AAGTCCCAGAGAAAGACCAGCAGGCCAAGAGAGCCAGCGCAAGGTG----- 375  
| : : : : :  
Db 121 SerGLyInthrLySGluLeuGINLaLyLySLyAsnGIuProThrValIThrProSer 140

QY 376 -----AGCCAGAACTACCCCATCGTGCACAACCTGAGGGGCCAAGTGTGCCACCGGCC 429  
| : : : : :  
Db 141 GLyGLYSerArGrAsnLyProIleVal-----SerValAsnAnGIuTrpValIHISGIuPro 159

QY 430 ATCAGCCCCCGCAACCTCGAAGCGCTGGGTGAAGTGTTCAGAGAGAAAGCCTTCAGGCC 489  
| : : : : :  
Db 160 LeuSerProArGrHrLeuAsnAlaITrpValLyValIleGIuGLuLySLySPheSerAla 179

QY 490 GAGGTGATCCCAATTGTTCAACCGCGCTGAGGAGAGGCGCCACCCCGACGAGACTGAAACAG 549  
| : : : : :  
Db 180 GluValValProMetCPheSerAlaLeuAlaGLuGLyVALIleProLyArAsnIleAsnGIu 199

QY	550	ATGTTGAAACCGGTGGGGCGGACCAAGCGCCGACATGCTGAAAGACACATCAAC	603
Db	200	MetLeuAsnAlaValGlyGluHisGlnGlyAlaLeuGlnIleValAllysAspValIleAsn	219
QY	610	GAGAGCGCCGCGAAGTGGACCGCTGTGCACCCCGTGTACAGCCGCGCCCGTGTCCCGCG	669
Db	220	GluGluAlaAlaAspTrpAspLeuArgHisProProGlnGlnIleProProAlaGlnGly	239
QY	670	CAGATGCGGAGACCCCGCGGAGCGACATCGCCCGCGCCACACAGACCCCTGTGAGAGCG	729
Db	240	ValLeuAspAspProGlnGlySerAspIleAlaGlyThrThrsThrIleGlnGluGln	259
QY	730	ATCGCCTGTGATACC--AGCAACCCCGCCGTGGCGCCGCGGAGCATTCACAAGCGGTGG	786
Db	260	IleGluTrpThrThrArgAlaGlnAsnAlaValAsnValGlyAsnIleTyrIysGlyTrp	279
QY	787	ATCATCTCGGCGCTGTAACAAGATCGTCGAGATGTCACGCCCTGTGACATCTTGGACATC	846
Db	280	IleIleLeuGlyLeuGlnHisCysValIlySmetTyrAsnProValAsnIleLeuAspIle	299
QY	847	CGCGAGGCGCCCAAGAGCGCTTCCGCGACTCGAGAACCGCTTCTTCAAGCCCGCGCG	906
Db	300	IysGlnGlyProIysGluProPheIysAspTyrValAspArgPheTyrIysAlaLeuArg	319
QY	907	GCCGAGCAGGCCACCCAGAGCGTGAAGAACTGATGATGACCCGACCCCTGTGTGCAAGAC	966
Db	320	AlaGluGlnThrAspProAlaValIlyAsnTrpMetThrGlnSerLeuLeuIleGlnAsn	339
QY	967	GCCAAACCCCGACTGCAAGACCATCTGTGGCGGCTCGCGGCCCGCGGACACCTGTGAGAG	1022
Db	340	AlaAsnProAspCysIlyThrValLeuIysGlyLeuGlyMetAsnProThrIleuGlnI	359
QY	1027	ATGATGACCGCGCTTGCAGAGCGGTGGCGGCGCCCGGACCAAGAGCCCGCGTGTGGCGAG	1086
Db	360	MetIleuThrAlaCysGlnGlyIlyIleGlyAlaGlnHisIlyValAlaArgLeuMetAlaGlu	379
QY	1087	GCGATGAC-----CAGGCCCAACAGCGTGAACATTCATGATGCAAG--AGC	1133
Db	380	AlaMetThrAlaAlaPheGlnGlnGlnIleThrValGlyAsnIlePheValGlnGlnIlyAla	399
QY	1132	AACTTCAAGGGCGCCC-----CGGCGCAACGTCAAGTGCCTC	1167
Db	400	ArgProArgGlyProLeuGlyGlyIlyArgGlyIlyArgProLeuAsnProAsnIleIysGlyTr	419
QY	1168	AACCTGCGGCAAGAGGCGCCACATGCGCCCAAGAACTCGCGCGCCCGCCGCAAGAGGCGTCC	1222
Db	420	AsnCysGlyIysProGlyHisLeuAlaArgPheCysIlyValAlaProIlyArgGlnIlyCys	439
QY	1228	TGGAAGTGCAGGCAAGAGGCGCACACAGATGTGAAGACTCAGACCGGCGGAGGCCAACTTC	1287
Db	440	TrpIysCysGlySerProAspHisGlnMetIysAspCys--GlnIysGlnIlyAsnIlePhe	458
QY	1288	CTGGGCAAGATCTGGCCAGCCACAGAGGCGCGCCCGGCAACTTCTGTGCAAGCCGACGC	1347
Db	459	LeuGly--PheGlyProTrpGlyIlyArgGlyIysProArgAsnIlePhe-----	472
QY	1348	GAGCGCGCGCGCCACCGTCCACACCGCGCCCGCGCGCGAGAG-----TTCCGCTTC	1401
Db	473	ProIleuThrSerIleArgProThrAlaProPheMetGluAlaGlyAspTyrSerArgPro	491
QY	1402	GAGAGAGC-----ACCGCGCGCGCC-----	1422
Db	492	GluGluAsnTrpTyrAlaAspArgProProThrArgIlyProGlyIysProAspArgProAla	511
QY	1423	-----AGCGAGAG-----	1433
Db	512	ThrAlaLeuLeuIysGlnTyrAlaValGlnGlyIysArgGlnIlyGlnGlnTrpGlnAsn	531
QY	1432	---CCCAAGACCGCGAGCGCTTACCGGAGCGCCCTGACCGCGCGCGAGCGCTGTGCAG	1488
Db	532	HisSerProGlnGlnSerProTyrGluGlnAlaTyrIserSerLeuAspGlySerLeuPheIly	551

## RESULT 15

FOLJTM

gag polyprotein - simian immunodeficiency virus (strain stm)

N:Alternate names: core polyprotein

N:Contains: core protein p17; core protein p24

C:Species: simian immunodeficiency virus, SIV

A:Note: host Macaca arcuoides (setup-tailed macaque)

A:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #exec\_change 09-Jul-2004

R:Khan, A.S.; Galvin, T.A.; Lowenstein, L.J.; Jennings, M.B.; Gardner, M.B.; Buckler, C.

J. Virol. 65, 7061-7065, 1991

A:Title: A highly divergent simian immunodeficiency virus (SIVstm) recovered from stored

A:Reference number: A41565; MUID:92046379; PMID:1942258

A:Accession: A41565

A:Molecule type: DNA

A:Reel/dues: 1-510 &lt;KHA&gt;

A:Cross-references: UNIPROT:P31634; GB:X60667; NID:G60527; PIDN:CAA43084.1; PID:G60528

C:Genetics:

A:Gene: gag

C:Superfamily: AIDS-related virus gag polyprotein

C:Keywords: core protein; polyprotein

F:1-115/Product: core protein p17 #status predicted &lt;P17&gt;

F:116-510/Product: core protein p24 #status predicted &lt;P24&gt;

## Alignment Scores:

Pred. No.:	2 81e-63	Length:	510
Score:	1396.00	Matches:	279
Percent Similarity:	67.88%	Conservative:	76
Best Local Similarity:	53.35%	Mismatches:	128
Query Match:	1	Indels:	40
DB:	1	Gaps:	11

US-09-475-704a-4 (1-1509) x FOLJTM (1-510)

```
QY 1 ATGGGCGCCGCGCCGAGCATCTCTGCGCGCGGAGAGCTGAGACAGTGGAGAGATCCGC 60
DB 1 MetGlyAlaArgSerSerValLeuSerGlyLysValAlaSpolLeuGlnLysValArg 20
QY 61 CTGGCGCCGCGCGCGAGAAAGCACTCATCTGAGACACCTGCTGGCGCGCGAG 120
DB 21 LeuArgProGlyGlyLysValLysValMetLeuLysValValAlaValAlaValAnglu 40
QY 121 CTGGAGGCGCTTGCCTGAAACCCGCGCTGCGAGACCGCGAGGCGCTGAGACATC 180
DB 41 LeuArgPheGlyLeuValAlaLysSerLysLeuGlnSerLysGlnLysValLysVal 60
QY 181 ATGAAGCAGCTGCAAGCCCGCGCTGCGAGACCGCGAGAGCTGCGAGCTGTACAC 240
DB 61 LeuThrValLeuGlnProLeuValProThrGlySerGlnLysLeuLysSerLeuPheAn 80
QY 241 ACCGTGGCCACCTGTACTGCTGCGAGCGCGGCGATCGAGCTCCGAGACACCAAGAGCC 300
DB 81 ThrValCysValLysValLysValLysValLysValLysValLysValLysValLysVal 100
QY 301 CTGGAGCAAGATGAGAGAGAG-----CAGAAACAAGTCCAGAG 339
DB 101 LysGlnValValLysValLysValLysValLysValLysValLysValLysValLysVal 120
QY 340 AAGACCCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 399
DB 121 ThrSerArgProThrAlaProProSerGlyArg---GlyGlyAsnTyrPro---ValGln 138
QY 400 AACCTGCAAGGCGCAAGTGTGACACAGGCGATCAGCCCGCGACCTTGAAGCGCTGAGT 459
DB 139 GlnValGlyGlyAsnTyrValLysValLysValLysValLysValLysValLysValLysVal 158
QY 460 AAGGTATGAGAGAGAGAGGCGCTTCAAGCCCGGAGGTATCCCATGTTTCCCGCCGAGC 519
DB 159 LysLeuValGlnGlnLysValLysValLysValLysValLysValLysValLysValLysVal 178
QY 520 GAGGCGCCACCCCGCAGCAGCTGAAACAGATGTTGAAACCGGTGGCGCGCCAGCGCC 579
DB 179 GlnGlyCysThrProTyrArgPheLeuGlnMetLeuAsnCysValGlyLysValLysVal 198
```

```
QY 580 GCCATGAGATGCTGAAGACACCATCAACGAGAGGCGCGGAGTGGAGCGGCTGCAC 639
DB 199 AlaMetGlnLeuLysArgGlnLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 218
QY 640 CCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATC 699
DB 219 Pro---GlnProGlyProLeuProAlaGlyGlnLeuArgGlnProSerGlySerLeuPhe 237
QY 700 GCGGCGCCACGAGACACCTGCAAGAGAGATGCGCTGATG---ACGACAAACCCCGCC 756
DB 238 AlaGlyThrThrSerThrValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 257
QY 757 GTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 816
DB 258 IleProValGlyAsnLysValLysValLysValLysValLysValLysValLysValLysVal 277
QY 817 ATGTACAGCCCGTGTAGCATCTGTGACATCCGCGAGGCGCGCGCGCGCGCGCGCGCGCG 876
DB 278 MetTyrAsnProValAsnLysLeuSerLysValLysValLysValLysValLysValLysVal 297
QY 877 TACGTGACCGCTTTCATGACACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 936
DB 298 TyrValArgPheTyrLysSerLeuArgLysGlnGlnAlaAspProAlaValLysAsn 317
QY 937 TGGATGACCGAGACCTCTGTGTGAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 996
DB 318 TrpMetLysGlnThrProLeuLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 337
QY 997 GCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1056
DB 338 GlyLeuGlyMetAsnProThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 357
QY 1057 CCGGCGCAAGAGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1116
DB 358 ProGlyGlnLysValArgLeuMetAlaGlnValAlaLysValLysValLysValLysValLysVal 377
QY 1117 ATGATGCAAGAGAGCACTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1176
DB 378 LeuProPheAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 397
QY 1177 AAGGAGGCGCATGCGCAAGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1236
DB 398 LysGlnGlnLysThrAlaLysGlnLysValLysValLysValLysValLysValLysValLysVal 417
QY 1237 GCGAAGAGGCGCACCAATGAAAGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1296
DB 418 GlyLysProGlyLysGlnMetAlaLysCysProGlyLysGlnValGlyPheLeuGly--- 436
QY 1297 ATCTGCGCCAGCACAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1341
DB 437 PheGlyProTyrGlyLys---LysProArgAsnPheProMetAlaGlnLysProGlnGly 455
QY 1342 GCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1401
DB 456 LeuThrProThrAlaProProGlnMetProThrAlaProProValAsp----- 471
QY 1402 GAGGAGACACCCCGCGC-----CAGAAACAAGTCCAGAG 1428
DB 472 -----ProAlaAlaLeuLeuArgSerTyrMetGlnLeuGlyLysValGln 487
QY 1429 GAGCCCAAGAGAGCGGAGCGCTTACCGC-----GAGCCCTTACCGCGCGCGCGCGCG 1479
DB 488 ArgGlnSerArgLysThrProTyrLysValLysValLysValLysValLysValLysValLysVal 507
QY 1480 CTGTTGCGC 1488
DB 508 LeuPheGly 510
```

Search completed: March 11, 2005, 15:43:51  
Job time : 103.775 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:39:57 ; Search time 174.728 Seconds  
(without alignments)  
5584.002 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 2772  
Sequence: 1 atggcgccgcgcgcagcat.....acgaccctcgcgcagctaa 1479

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1396920 seqs, 329844858 residues

Total number of hits satisfying chosen parameters: 2793840

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+np.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_epool\_p/US09475704/runat\_10032005\_140224\_14937/app\_query.fasta\_1.3342  
-DB=Published Applications AA -QFMT=fastran -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEATSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09475704.OCN\_1.1\_466 @runat\_10032005\_140224\_14937  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DBPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/1/pubpaa/US6\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/1/pubpaa/US6\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	2606	94.0	492	10	US-09-899-575-17	Sequence 17, Appl
2	2448	88.3	492	16	US-10-332-413-2	Sequence 2, Appl
3	2435	87.8	492	9	US-09-991-258-5	Sequence 5, Appl
4	2435	87.8	631	16	US-10-333-413-8	Sequence 8, Appl
5	2423	87.4	492	14	US-10-339-217-110	Sequence 110, Appl
6	2416	87.2	492	14	US-10-339-217-109	Sequence 109, Appl
7	2415.5	87.1	491	14	US-10-339-217-140	Sequence 140, Appl
8	2407.5	86.9	491	14	US-10-339-217-104	Sequence 104, Appl
9	2396.5	86.5	491	14	US-10-339-217-143	Sequence 143, Appl
10	2382	85.9	496	14	US-10-339-217-108	Sequence 108, Appl
11	2381.5	85.9	487	14	US-10-339-217-106	Sequence 106, Appl
12	2381.5	85.9	497	14	US-10-339-217-107	Sequence 107, Appl
13	2344	84.6	508	14	US-10-339-217-105	Sequence 105, Appl
14	2343	84.5	502	10	US-09-899-575-22	Sequence 22, Appl
15	2328.5	84.0	494	14	US-10-339-217-126	Sequence 126, Appl
16	2297	82.9	500	14	US-10-339-217-18	Sequence 18, Appl
17	2277	82.1	498	16	US-10-325-468-22	Sequence 22, Appl
18	2266	81.7	500	16	US-10-325-468-8	Sequence 8, Appl
19	2263	81.6	500	16	US-10-325-468-34	Sequence 34, Appl
20	2262	81.6	498	16	US-10-325-468-27	Sequence 27, Appl
21	2258	81.5	500	14	US-10-339-217-121	Sequence 121, Appl
22	2256	81.4	500	14	US-10-097-534-32	Sequence 32, Appl
23	2256	81.4	500	14	US-10-059-271-85	Sequence 85, Appl
24	2256	81.4	500	17	US-10-102-622-4	Sequence 4, Appl
25	2256	81.4	500	17	US-10-844-658-6	Sequence 6, Appl
26	2252.5	81.3	499	15	US-10-296-734-1	Sequence 1, Appl
27	2252	81.2	500	16	US-10-325-468-17	Sequence 17, Appl
28	2251	81.2	500	9	US-09-968-355-26	Sequence 26, Appl
29	2251	81.2	583	9	US-09-968-355-17	Sequence 17, Appl
30	2249	81.1	500	14	US-10-223-172A-42	Sequence 42, Appl
31	2249	81.1	500	14	US-10-224-999A-3483	Sequence 3483, Ap
32	2246.5	81.0	501	14	US-10-339-217-136	Sequence 136, Appl
33	2245	81.0	1350	10	US-09-952-060-35	Sequence 35, Appl
34	2245	81.0	1550	16	US-10-380-641-35	Sequence 35, Appl
35	2237	80.7	557	14	US-10-339-217-144	Sequence 144, Appl
36	2233	80.6	500	14	US-10-339-217-122	Sequence 122, Appl
37	2226.5	80.3	503	14	US-10-339-217-127	Sequence 127, Appl
38	2225	80.3	493	14	US-10-339-217-111	Sequence 111, Appl
39	2221.5	80.1	499	17	US-10-844-658-2	Sequence 2, Appl
40	2213.5	79.9	494	14	US-10-339-217-100	Sequence 100, Appl
41	2212.5	79.8	513	14	US-10-339-217-31	Sequence 31, Appl
42	2211.5	79.8	494	14	US-10-339-217-99	Sequence 99, Appl
43	2204.5	79.5	494	14	US-10-339-217-101	Sequence 101, Appl
44	2199.5	79.3	496	14	US-10-339-217-32	Sequence 32, Appl
45	2195	79.2	512	16	US-10-093-953A-25	Sequence 25, Appl

## ALIGNMENTS

RESULT 1  
US-09-899-575-17  
; Sequence 17, Application US/09899575  
; Publication No. US20030223961A1  
GENERAL INFORMATION:  
; APPLICANT: Zur Negele, Jan  
; APPLICANT: Barnette, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Batrelita Janse  
TITLE OF INVENTION: POLYPEPTIDES ENCODING ANTIGENIC HIV TYPE C  
FILE REFERENCE: PP01631.102  
CURRENT APPLICATION NUMBER: US/09/899,575  
PRIOR FILING DATE: 2001-07-05  
PRIOR APPLICATION NUMBER: 09/475,704  
NUMBER OF SEQ ID NOS: 135  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 17  
LENGTH: 492  
TYPE: PRT  
ORGANISM: Human immunodeficiency virus

US-09-899-575-17

## Alignment Scores:

Pred. No.: 2,666-129 Length: 492  
 Score: 2606.00 Matches: 492  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 94.01% Indels: 0  
 DB: 10 Gaps: 0

US-09-475-704a-3 (1-1479) x US-09-899-575-17 (1-492)

```

QY 1 ATGGGCGCCGCGCAGCATCTCTGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCG 60
DB 1 MetClyAlaArgAlaSerIleLeuArgGlyGlyLeuAspAlaTrpGluArg 20
QY 61 CTGGCGCCCGCGCGCAAGAGTGTACATGATGAGCACTGTGTGGCCAGCCGCGAG 120
DB 21 LeuArgProGlyGlyLysCysTyrMetMetLysIleValTrpAlaSerArgGlu 40
QY 121 CTGGAGAGATTCCGCTGAGACCCCGCTGCTGTGAGACCGAGGCTGCAAGCATC 180
DB 41 LeuGluLysPheAlaLeuAsnProGlyLeuLeuGluThrSerGluGlyCysLysGlnIle 60
QY 181 ATCCGCGAGCTGACCCCGCTGACAGCCGCGAGCGAGAGCTGAAGAGCTGTTCAC 240
DB 61 IleArgGlnLeuHisProAlaLeuGlnThrGlySerGluLeuLysSerLeuPheAsn 80
QY 241 ACCGTGGCCACCTGTACTGCTGTGACAGAGAGATCGAGTCCGCGACACCAAGAGCC 300
DB 81 ThrValAlaThrLeuTyrCysValHisGluLysIleGluValArgAspThrLysGluAla 100
QY 301 CTGGAGAGAGTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 101 LeuAspLysIleGluGluGlnGlnAsnLysCysGlnGlnLysIleGlnGlnAlaGluAla 120
QY 361 GCCGAGAGAGGAGAGGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 121 AlaAspLysGlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 140
QY 421 GTGACACAGGCGCATCAGCCCGCGACCTTGAACGCTGTGGGTGAAGTGAAGAGAG 480
DB 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLys 160
QY 481 GCCTTCAGCCCGGAGGTATCCCGCATGTTCAACGCGCTGAGAGAGAGAGAGAGAG 540
DB 161 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 180
QY 541 GACCTGAACAGATGTTGAACACCGTGGCGGCGCACAGCGCCATGACAGATGTAAG 600
DB 181 AspLeuAsnThrMetLeuAsnThrValGlyHisGlnAlaAlaMetGlnMetLeuLys 200
QY 601 GACACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 201 AspThrIleAsnGlnGlnAlaAlaGlnTrpAspArgValHisProValHisAlaGlyPro 220
QY 661 ATCGCCCGCGCGCAGATGCGAGAGCCCGCGGAGAGAGATGCGCGCGCGAGAGAG 720
DB 221 IleAlaProGlyGlnMetCysArgLysProArgGlySerAspIleAlaGlyThrThrSerThr 240
QY 721 CTGGAGAGAGAGATGCGCTGTGATGACAGCAACCCCGCATCCCGTGGGAGCATTA 780
DB 241 LeuGlnGlnGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyr 260
QY 781 AAGCGGTGATCATCTGGGCTGGAACAAGATCTGCGGATGTAAGAGAGAGAGAGAG 840
DB 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle 280
QY 841 CTGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 281 LeuAspIleLysGlnGlyProLysGlnTrpPheAlaGlyAspTyrValHisArgPhePheLys 300
QY 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
  
```

```

DB 301 ThrLeuArgAlaGluInserThrGlnGlnValLysAsnTrpMetThrAspThrLeuLeu 320
QY 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340
QY 1021 CTGGAGAGAGATGATGACCGCTGCGCAGGCGGTGGGCGGCCAGCCAGAGCCGCTG 1080
DB 341 LeuGluGlnMetMetThrAlaCysGlnGlnValGlyGlyProSerHisLysAlaArgVal 360
QY 1081 CTGGCGGAGAGAGATGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 361 LeuAlaGlnAlaMetSerGlnAlaAsnThrSerValMetMetGlnLysSerAsnPheLys 380
QY 1141 GGCAGCGCGCGCATCGTCAAGTGTCTTCACTGCGGAGAGAGAGAGAGAGAGAGAG 1200
DB 381 GlyProArgArgIleValLysCysPheAsnGlyGlyLysGlnGlyHisIleAlaArgAsn 400
QY 1201 TGCAGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 401 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisGlnMetLys 420
QY 1261 GACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 421 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCCGAGCACTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 441 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGlnSerPheArg 460
QY 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 461 PheGluGlnThrThrProGlyGlnLysGlnLysSerLysAspArgGluThrLeuThrSer 480
QY 1441 CTGAAGAGCGTGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
DB 481 LeuLysSerLeuPheGlyAsnAspProLeuSerGln 492
  
```

## RESULT 2

```

; Sequence 2, Application US/10332413
; Publication No. US20040116660A1
; GENERAL INFORMATION:
; APPLICANT: Johnstone, Robert Edward
; APPLICANT: Swannstrom, Ronald Ivar
; APPLICANT: Morris, Lynn
; APPLICANT: Karim, Salim Abdool
; APPLICANT: Williamson, Carolynn
; TITLE OF INVENTION: Processes for the selection of HIV-1 subtype C isolates, selected H
; TITLE OF INVENTION: Subtype Isolates, Their Genes and Modifications and Derivatives T
; FILE REFERENCE: 45669-281993
; CURRENT APPLICATION NUMBER: US/10/332,413
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: PCT/IB01/01208
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/216,995
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: ZA 2000/3437
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: ZA 2000/4924
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; US-10-332-413-2
  
```

## Alignment Scores:

Pred. No.: 5,56-121 Length: 492  
 Score: 2448.00 Matches: 461

Percent Similarity: 97.36% Conservative: 19  
 Best Local Similarity: 93.51% Mismatches: 11  
 Query Match: 88.31% Indels: 2  
 DB: 16 Gaps: 2

US-09-475-704a-3 (1-1479) x US-10-332-413-2 (1-492)

```

QY 1 ATGGGGGCGCCGCGCAAGATCCCTGGCGCGCGCAAGCTGAGCCCTGGAGCGCCATCCGC 60
DB 1 MetcGlyAlaArgAlaSerIleLeuArgGlyGluLeuAspArgLeuGlyTrpGluValLeuArg 20
QY 61 CTGGCGCGCGCGCGCAAGAGTCTACATGATGAACACCTGGTGTGGCGCAAGCGCGAG 120
DB 21 LeuArgProGlyGlyValSerIleValMetLeuValIleValIleValIleValIleValIle 40
QY 121 CTGAGAGAGTTGGCTGAAACCCCGGCTGTGGAGACCAAGCGGCTGGCAAGCATC 180
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlyGlyValSerGlyGlyIle 60
QY 181 ATCCCGGCGCTGCAACCCCGCTGGAGACCGGCGAGAGAGTGAAGAGCTGTTCAC 240
DB 61 MetLeuGluLeuGluProAlaLeuGluThrGlyThrGluGluLeuValSerLeuValSerLeu 80
QY 241 ACCGTGGCGCAACCTGTACTGCGTGACAGAGAGATGAGAGTCCGCGCAACCAAGAGGCG 300
DB 81 ThrValAlaThrLeuValTrpValIleGluValIleGluValIleGluValIleGluValIle 100
QY 301 CTGACCAAGATCGAGAGAGAGCAAGATGCGCAGAGAGATCCAGAGCGCGCGAGGCG 360
DB 101 LeuAspValIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120
QY 361 GCGGCAAGAGGCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 121 AlaAsp--GlyValValSerGluAsnValProIleValGluAsnLeuGluGluGluGluGlu 139
QY 421 GTGCAACAGAGCGCATGAGCGCGCGCAACCTGGAAGAGTGTGAGAGAGAGAGAGAGAG 480
DB 140 ValIleGluAlaIleSerProAlaGluLeuAsnAlaIleTrpValIleValIleGluGluGlu 159
QY 481 GCGTTCAGCGCGCGAGGTGATCCCATGTTCAACCGCTTGAAGAGGCGCGCGCGAGCG 540
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGlyGluValIleAlaThrProGlu 179
QY 541 GACCTGGAACAGATGTTGAACAACCGTGGCGCGCAACAGCGCGCGCATGAGATGCTGAAG 600
DB 180 AspLeuAsnThrMetLeuAsnThrValIleGlyGlyValIleGluAlaIleValIleValIle 199
QY 601 GACACCATCAAGAGAGCGCGCGAGTGGAGCGCGGTGACACCGCGTGAACGCGCGCGCG 660
DB 200 AspThrIleAsnGluGluValIleAlaGluTrpAspArgValIleProValIleAlaGlyPro 219
QY 661 ATCGCGCGCGCGAGATGCGCGAGCGCGCGCGAGAGATCGCGCGCGCAACCAAGCAGC 720
DB 220 IleAlaProGlyGluMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239
QY 721 CTGCGAGAGCAAGATGCTGTGATGACAGCAACCGCGCGCATCCCGTGGCGCGCATCTAC 780
DB 240 LeuGluGluGluGluIleAlaThrMetThrSerAsnProIleProValIleValIleValIle 259
QY 781 AAGCGGTGATCATCTGGGCGTGAACAAGATCGTGGATGAGAGAGCGCGGTGAGCATC 840
DB 260 LysArgTrpIleIleLeuGlyLeuAsnValIleValIleValIleValIleValIleValIle 279
QY 841 CTGACATCAAGAGAGCGCGCGCAAGAGCGCGCTTCCGAGTACGTGAGCGCGCTTCTTCAAG 900
DB 280 LeuAspIleArgGluGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299
QY 901 ACCCTGGCGCGCGAGAGAGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 300 ThrLeuArgAlaGluGluIleAlaThrGluGluValIleValIleValIleValIleValIle 979
QY 961 GTGCAAGAGCGCAACCGCGAGTGAAGAGCATCTCGCGGCTCTGGCGCGCGCGCGCAGC 1020
DB 981 GTGCAAGAGCGCAACCGCGAGTGAAGAGCATCTCGCGGCTCTGGCGCGCGCGCGCAGC 1040

```

```

DB 320 ValGluAsnAlaAsnProAspCysIleThrIleLeuArgAlaLeuGlyProGlyAlaThr 339
QY 1021 CTGAGAGAGATGATGACCGCTTGGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
DB 340 LeuGluGluMetMetThrIleValIleValIleValIleValIleValIleValIleValIle 359
QY 1081 CTGGCGCGAGGCGATGAGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137
DB 360 LeuAlaGluAlaMetSerGluIleValIleValIleValIleValIleValIleValIleValIle 379
QY 1138 AAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1197
DB 380 LysGlyProArgArgIleValIleValIleValIleValIleValIleValIleValIleValIle 399
QY 1198 AACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1257
DB 400 AsnValArgAlaProArgIleValIleValIleValIleValIleValIleValIleValIle 419
QY 1258 AAGGAGTGAAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1317
DB 420 LysAspCysThrGluArgGluIleAlaAsnPheLeuGlyLysIleTrpProSerIleValIle 439
QY 1318 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377
DB 440 ArgProGlyAlaAsnPheLeuGluIleAlaAsnArgProGluProThrAlaProProAlaGluSerPhe 459
QY 1378 CGCTTGAAGAGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437
DB 460 ArgPheGluGluThrThrProAlaProLysGluGluProIleGluArgGluProLeuThr 479
QY 1438 AGCCTGAAGAGCGCTGTGGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1496
DB 480 SerLeuLysSerLeuPheIleValIleValIleValIleValIleValIleValIleValIle 492

RESULT 3
US-09-991-258-5
; Sequence 5, Application US/09991258
; Patent No. US2002014975A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Robert
; APPLICANT: Kelch, Paula
; APPLICANT: Dryga, Sergey
; APPLICANT: Caley, Ian
; APPLICANT: Maughan, Maureen
; APPLICANT: Johnston, Robert
; APPLICANT: Davis, Nancy
; APPLICANT: Swanson, Ronald
; TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOES WITH MODIFIED HIV GENES FOR USE
; FILE REFERENCE: 01113.0001U3
; CURRENT APPLICATION NUMBER: US/09/991,258
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 09/902,537
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,995
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence, No. US2002014975A1e =
US-09-991-258-5

Alignment Scores:
Pred. No.: 2,666-120 Length: 492
Score: 2435.00 Matches: 458
Percent Similarity: 96.96% Conservative: 20
Best Local Similarity: 92.90% Mismatches: 13
Query Match: 87.84% Indels: 2

```





```
Db 31 MetAlaIaIaArgHisSerIleLeuArgGlyGluIuLyLeuAspIyStrpGluIySileArg 50
QY 61 CTGGCGCCCGCGGCAAGAGTGCTCATGATGAGCAACCTGGTGGGCGCCAGCCCGGAG 120
Db 51 LeuArgProGlyGlyIySlySbIyStrpMetLeuIyHisIleValITrpAlaSerArgGlu 70
QY 121 CTGGAGAAATTGGCCCTGAAACCCCGGCTGCTGGAGAACCGGAGGCTGGCAAGCATC 180
Db 71 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluIyCysIySgIinIle 90
QY 181 ATCCGCGACCTGACACCCCGCTGACAGCCGCGAGCGAGGAGTGAAGAGCTGTCAAC 240
Db 91 MetIySgIinLeuGluInProAlaLeuGluInThrGlyThiGluIuLeuIySerLeuIyTrpAsn 110
QY 241 ACCGTGGCCACCTGTACTGCTGTCAGAGAGATGAGGTCCGCGACCAACAGAGAGCC 300
Db 111 ThrValAlaIaThrLeuIyrcySValHisGluIyIleGluValArgAspThrIySgIuAla 130
QY 301 CTGGACAAATGCGAGAGAGAGAGCAAGTGCACAGAAAGATCCAGCGAGCGCGAGCC 360
Db 131 LeuAspIySbIleGluIuGluGluInAsnIySbGluInIySbThrGluInIuAlaIySbAla 150
QY 361 GCCGACAAAGGCGAGGTGAGCCAGAACTACCCCATGCTGTCAGAACCTGACAGGCGCATG 420
Db 151 AlaAsp---GlyIySbValSerGluInAsnIyrcProIleValGluInAsnLeuGluInIySbMet 169
QY 421 GTGCACAGAGCCCATGACCCCGCGACCTTGACGCTGGTGAAGGTGATGAGAGAGAG 480
Db 170 ValHisGluInAlaIleSerProArgTrpThrLeuAsnAlaItrpValIySbValIleGluIuIySb 189
QY 481 GCCTTACAGCCCGAGGTGATCCCATGTTTCAACCGCTTGAAGCGAGGCGCGACCCCGGAG 540
Db 190 AlaPheSerProGluIuAlaIleProMetPheThrAlaLeuSerIuGluIyAlaItrpProGluIn 209
QY 541 GACCTGAACACGATGTTGAAACACCTGGGCGGCGACCAAGGCGCGCATGACATGCTGAAG 600
Db 210 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGluInAlaIaMetGluInMetLeuIySb 229
QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCGTGCACCCCGCGGAGCCCGG 660
Db 230 AspThrIleAsnGluIuGluInAlaIaGluIuTrpAspArgLeuHisProValHisAlaGlyPro 249
QY 661 ATGCGCCCGCGGCAATGCGCGAGCCCGCGGAGAGAGCATGCGCGGACCAACAGCAGCACC 720
Db 250 IleAlaProGlyGluInMetArgIuProArgGlySerAspIleAlaGlyThrTrpSerThr 269
QY 721 CTGGAGAGACAGATGCGCTGATGATGACAGCAACCCCGCATCCCGTGGGCGCATCTTAC 780
Db 270 LeuGluInGluInIleAlaItrpMetTrpSerAsnProProIleProValGlyAspIleIyrc 289
QY 781 AAGCGGTGATCATCTGTGGGCTTGAAACAAGATGTCGAGATGTACAGCCCGGTGAGCATC 840
Db 290 IySbArgStrpIleIleLeuGlyLeuAsnIySbIleValArgMetIyrcSerProValSerIle 309
QY 841 CTGGACATCAAGAGAGGCGCGCAAGAGCCCTTCCGAGCATGCTGAGACCGCTTCTTCAAG 900
Db 310 LeuAspIleArgGluInIyrcProIySbIuProPheArgAspIyrcValAspArgPhePheIySb 329
QY 901 ACCCTGCGGCGCGAGAGAGCAACCGAGAGGTGAAGAACTGATGACCGACCTGTGCTG 960
Db 330 ThrLeuArgAlaGluInAlaIaThrGluInGluValIyAsnAsnTrpMetThrAspThrLeuLeu 349
QY 961 GTGCAGAACGCGCAACCCCGCATGCAAGACCATCTGCGGCTGTGGCGCCCGCGCGCAGC 1020
Db 350 ValGluInAsnAlaAsnProAspCysIySbThrIleLeuArgAlaLeuGlyProGlyAlaItrp 369
QY 1021 CTGGAGAGAAATGTAACCGCTGCGAGGCGTGGGCGGCGCGCGAGCGCAAGAGCCCGCTG 1080
Db 370 LeuGluInMetMetThrAlaCysGluInGlyValGlyIyrcProGlyIySbIySbAlaArgVal 389
QY 1081 CTGGCGAGGCGCATGAGCCAGCGCAACAC---AGCGTATGATGAGAGAGCAAGCACTTC 1137
```

```
Db 390 LeuAlaGluAlaMetSerGluInThrAsnSerGlyAsnIleMetMetGluInArgSerAsnPhe 409
QY 1138 AAGGCGCCCGCGCGCATCTGTCAGATGCTTCAATCTGGCGCAAGAGCGCAACCGCGCGC 1197
Db 410 IySbGlyProArgArgIleValIySbCysPheAsnCysGlyIySbGluIyHisIleAlaArg 429
QY 1198 AACTGCGCGCGCGCGCAAGAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCACAGATG 1257
Db 430 AsnCysArgAlaProArgIySbIySbGlyCysTrpIySbCysGlyIySbGluIyHisGluInMet 449
QY 1258 AAGCATGCAACGAGCGCGCGCAACCTTCTGGCGCAAGATGCTGGCCAGCAACAGAGCG 1317
Db 450 IySbAspCysThrGluInArgGluInAlaAsnPheLeuGlyIySbIleTrpProSerHisIySbGly 469
QY 1318 CGCGCGCGCAACTTCTGTCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGAGCTTC 1377
Db 470 ArgProGlyAsnPheLeuInAsnArgProGluProThrAlaProProAlaGluSerPhe 489
QY 1378 CGCTTGCAGAGAGCGCAACCCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
Db 490 ArgPheGluInThrThrProAlaProIySbGluInProIleGluArgGluProLeuThr 509
QY 1438 AGCTGAGAGAGCTGTTCGCAACGACCCCTGAGCGCAG 1476
Db 510 SerLeuIySerLeuPheGlySerAspProLeuSerGlu 522
```

## RESULT 5

```
US-10-339-217-110
; Sequence 110, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xieohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Mengang
; TITLE OF INVENTION: Method for Identifying Compounds Which Inhibit Binding of
; TITLE OF INVENTION: Nucleocapsid 7 Protein to HIV-1 RNA
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: PaedSeq for Windows Version 4.0
; SEQ ID NO 110
; LENGTH: 492
; TYPE: PR
; ORGANISM: HIV-1
US-10-339-217-110
```

## Alignment Scores:

```
Pred. No.: 1.14e-119 Length: 492
Score: 2423.00 Matches: 451
Percent Similarity: 96.96 Conservative: 27
Best Local Similarity: 91.48 Mismatches: 13
Query Match: 87.41 Indels: 2
DB: 14 Gaps: 2
```

US-09-475-704a-3 (1-1479) x US-10-339-217-110 (1-492)

```
QY 1 AAGGCGCGCGCGCGCATCTGCGCGCGCGCGCAAGCTTGAGCGCGCATCCG 60
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGluIySbLeuAspThrTrpGluIySbIleArg 20
QY 61 CTGGCGCGCGCGCGCAAGAGTGCTCATGATGAGCAACCTGGTGGGCGCCAGCCCGGAG 120
Db 21 LeuArgProGlyGlyIySbIySbCysTrpMetLeuIyHisIleValITrpAlaSerArgGlu 40
QY 121 CTGGAGAAATTGGCCCTGAAACCCCGGCTGCTGGAGAACCGGAGGCTGGCAAGCATC 180
Db 41 LeuGluArgPheSerLeuAsnProGlyLeuLeuGluInThrSerGluIyCysIySgIinIle 60
```

QY 181 ATCCGCAAGCTGACCCCGCTGACAGACCGGAGGAGCTGAAGACCTGTTCAAC 240  
Db : : : : :  
QY 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlnIleuLysSerLeuTyraAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCTGTCACGAGAAAGATCGAGTCCGCGACACCAAGAGGCC 300  
Db : : : : :  
QY 81 ThrValAlaThrLeuPheCysValHisGlnLysIleAlaValArgSerThrLysGlnAla 100  
QY 301 CTGGACAAGATGAGAGGAGGAGGAGCAAGTGCACAGAAAGATCCAGAGGCGCCAGGCC 360  
Db : : : : :  
QY 101 LeuAspLysIleGlnGlnGlnIleuAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCCGACAAGGCGCAAGGTGAGCCGAACTACCCCATCGTGCGAGAACTGCAAGGCGCGATG 420  
Db : : : : :  
QY 121 AlaAsp---GlyThrValSerGlnAsnLysProIleValGlnAsnLeuGlnGlnMet 139  
QY 421 GTGGACCAAGGCGCATGACCCCGCGACCTTGAAAGCTGGGTGAGGTGATCGAGAGAA 480  
Db : : : : :  
QY 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLys 159  
QY 481 GCCTTCAGCCCGGAGGTGATCCCATGTTCAACGCGCTTGAAGAGGCGCCACCCCGCG 540  
Db : : : : :  
QY 160 AlaPheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlnAlaThrProGln 179  
QY 541 GACCTGACAAGATGTTGAACAACCTGCGCGCGCCACAGGCGCGCATGCGAGATGCTGAAG 600  
Db : : : : :  
QY 180 AspLeuAsnThrMetLeuAsnThrValGlnGlnHisGlnAlaAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAAGAGAGAGGCGCGCGAGTGGACCGCGGTGCACCCGCGCGCGCC 660  
Db : : : : :  
QY 200 AspThrIleAsnGlnIleGlnIleAlaGlnLysPheArgLeuHisProValHisAlaGlnPro 219  
QY 661 ATCGCCCGCGCGAGATGCGCGAGCGCGCGCGAGCATGCGCGCGCGCCACCGCGAGCC 720  
Db : : : : :  
QY 220 IleAlaProGlnGlnMetArgLysProArgLysSerAspIleAlaGlnLysThrSerThr 239  
QY 721 CTGGAGAGACAGATGCGCTGATGACCAACACCCCGCATCCCGCTGGCGCGACATCTAC 780  
Db : : : : :  
QY 240 LeuGlnGlnGlnIleAlaTrpMetThrAsnAsnProProValProValGlnLysIleTy 259  
QY 781 AAGCGGTGATCATCTGGGCTGGAACAGATCGTGCAGATGTAAGCCCGCTGAGCATC 840  
Db : : : : :  
QY 260 LysArgTrpIleIleLeuGlnLysLeuAsnLysIleValArgMetLysSerProValSerIle 279  
QY 841 CTGGACATCAAGAGAGGCGCGCGAGGCGCTCCGCGCATACGAGGACCGCTTCTTCAG 900  
Db : : : : :  
QY 280 LeuAspIleArgGlnIleProLysGlnProPheArgAspLysValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCGAGAGAGACCAAGAGGTGAAGAACTGTAGTACCGACACCTGCTG 960  
Db : : : : :  
QY 300 ThrLeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGGAGAACCGCAACCGCGCATGCGAAGACATCTCGCGCTCTGCGCCCGCGCGCGAC 1020  
Db : : : : :  
QY 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlnProIleValAlaThr 339  
QY 1021 CTGGAGAGATGATACCGCTGCGAGGCGGTGGCGCGCGCGCGCGACCAAGGCGCGG 1080  
Db : : : : :  
QY 340 LeuGlnGlnMetThrAlaCysGlnGlnValGlnGlnProSerHisLysAlaArgAl 359  
QY 1081 CTGGCGAGGCGATGAGCGCGCG---AAGACCAAGCTGATGATGACAGAGCAACTTC 1137  
Db : : : : :  
QY 360 LeuAlaGlnLysLysSerGlnThrAsnAsnAlaAsnIleMetGlnArgSerAsnHe 379  
QY 1138 AAGGCGCGCGCGCGATGCTCAAGTCTTCACTCGCGCGAGAGGCGCGACATCGCCCGC 1197  
Db : : : : :  
QY 380 LysGlnProArgArgIleIleLysCysPheAsnCysGlnLysGlnGlnHisLysAlaArg 399  
QY 1198 AACTCGCGCGCGCGCGCAAGAGGCTGCTGAGAGTGGCGCAAGAGGCGCGACAGATG 1257  
Db : : : : :  
QY 400 AsnGlnArgAlaProArgLysLysGlnLysCysGlnLysGlnGlnHisLysAlaArg 419  
QY 1258 AAGGACGTGACCGAGCGCGCAAGCTTCTCGGCGAAGATCTGGCGCGACCGCAAGGCGC 1317

Db : : : : :  
QY 420 LysAspCysThrGlnArgGlnAlaAsnPheLeuGlnLysIleTrpProSerHisLysGln 439  
QY 1318 CGCCCGCGCAACTTCTGTGAGAGCGCGCGCGAGCCACCGCGCGCGCGAGACTTC 1377  
Db : : : : :  
QY 440 ArgProGlnLysAsnPheLeuGlnAsnArgProGlnProThrAlaProProAlaGlnSerPhe 459  
QY 1378 CGCTTCGAGAGAACACACCCCGCGCGAGAGGAGAGACAGACCGCGAGACCTTGACC 1437  
Db : : : : :  
QY 460 ArgPheGlnGlnLysThrThrProAlaProLysGlnGlnProArgLysGlnLysProLeuThr 479  
QY 1438 AGCCTGAAGAGCCTGTGCGACAGACCGCGCTGAGCGCAG 1476  
Db : : : : :  
QY 480 SerLeuLysSerLeuPheGlnLysSerAspProLeuSerGln 492

RESULT 6  
US-10-339-217-109  
Sequence 109, Application US/10339217  
Publication No. US20030198648A1  
GENERAL INFORMATION:  
APPLICANT: Buecenter, Douglas  
APPLICANT: Hou, Xiaohong  
APPLICANT: Marlor, Christopher W.  
APPLICANT: Rice, William G.  
APPLICANT: Yang, Mengang  
TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of  
FILE REFERENCE: 111021.143 (ACH-US1)  
CURRENT APPLICATION NUMBER: US/10/339, 217  
CURRENT FILING DATE: 2003-01-09  
PRIOR APPLICATION NUMBER: US 60/347,369  
PRIOR FILING DATE: 2002-01-11  
NUMBER OF SEQ ID NOS: 144  
SOFTWARE: fastseq for Windows Version 4.0  
SEQ ID NO 109  
LENGTH: 492  
TYPE: PRT  
ORGANISM: HIV-1  
US-10-339-217-109

Alignment Scores:  
Pred. No.: 2,66e-119 Length: 492  
Score: 2416.00 Matches: 455  
Percent Similarity: 95.74% Conservative: 17  
Best Local Similarity: 92.29% Mismatches: 19  
Query Match: 87.16% Indels: 2  
DB: Gaps: 2

US-09-475-704A-3 (1-1479) x US-10-339-217-109 (1-492)

QY 1 ATGGGCGCGCGCGCGAGATCTTGGCGCGCGCAAGCTGAGCGCTGAGAGCGATCCGC 60  
Db : : : : :  
QY 1 MetGlnAlaSerAlaSerIleLeuArgGlnLysLysLeuAspLysTrpGlnLysAlaArg 20  
QY 61 CTGGCGCGCGCGCGCAAGAGTGTACTATGATGAAGACCTGTGTGGCGCGCGCGAG 120  
Db : : : : :  
QY 21 LeuArgProGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40  
QY 121 CTGGAAGATTCGCTGTAACCCCGCGCGCTGTGAGACCAAGGCGCTGCAAGCATC 180  
Db : : : : :  
QY 41 LeuGlnArgPheAlaLeuAsnSerGlnLeuLeuGlnLysThrAlaGlnGlnCysLysGlnIle 60  
QY 181 ATCCGCGAGCGTGCACCCCGCGCTGCGACACCGCGCGAGAGAGCTGGAAGACCTTCAAC 240  
Db : : : : :  
QY 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlnThrGlnLysLysSerLeuTyraAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCTGTCACGAGAAAGATCGAGTCCGCGACACCAAGAGGCC 300  
Db : : : : :  
QY 81 ThrValAlaThrLeuTyraCysValHisAlaGlnIleGlnValArgAspThrLysGlnAla 100  
QY 301 CTGGACAAGATGAGAGGAGGAGGAGCAAGTGCACAGAAAGATCCAGAGGCGCGAGGCC 360  
Db : : : : :  
QY 101 LeuAspLysIleGlnGlnGlnIleuAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120



Db 160 AlaPheSerProGluValIlePrometPheThrAlaLeuSerGluGlyAlaThrProGln 179  
QY 541 GACCTGAACGAGTGTGAAACACCGTGGCGGCGCCACGAGCCGCGCATGCGATCTGAAG 600  
Db 180 AspLeuAsnThrMetLeuSerThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLys 199  
QY 601 GACACCATCAACAGAGGCGCGCGAGTGGGACCGGCGTACCCGCTGACCGCGCGCC 660  
Db 200 AspThrIleAsnGluThrAlaIleGluThrPaspArgLeuHisProValHisAlaGlyPro 219  
QY 661 ATCGCCCGCGCGAGATGCGCGAGCCCGCGCGAGCGACATCGCCGCGACACGAGCACC 720  
Db 220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerSer 239  
QY 721 CTGGACAGACGATGCCCTTGATGACGACGACACCCCGCATCCCGCTGGCGCATCTAC 780  
Db 240 LeuGlnGluGlnIleAlaThrMetThrGlyAsnProProValProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATGCGGATGATGACGCGCGCGGAGCATC 840  
Db 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIle 279  
QY 841 CTGGACATCAAGAGCGCGCGCGAGAGCCCTTCCGAGTACGTGACCGCTTCTTCAAG 900  
Db 280 LeuAspIleLeuGlnGlyProLysGluProPheArgAspTyrValAspArgPheLeuLys 299  
QY 901 ACCCTGGCGCGAGAGAGCAACCCGAGAGTGAAGAATCTGATGACGACACCTGCTG 960  
Db 300 ThrLeuArgAlaGlnIleAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGCAAGAGCGCAACCCGAGTGAAGACCATCTCGCGCTCTGGCGCGCGCGCGCAGC 1020  
Db 320 ValGlnAsnAlaAsnProAspGlySerThrIleLeuArgAlaLeuGlyProGlyLysAsn 339  
QY 1021 CTGAGAGATGATGACCCGCTGCGAGCGCGGTGGCGCGCGCGCGCAAGCGCGCTG 1080  
Db 340 LeuGlnGluMetMetThrAlaCysGlnGlyValGlyProSerHisLysAlaArgVal 359  
QY 1081 CTGGCGAGCGAGTGAAGCGAGCGCACACGCGTGAATGATGACAGAGCAACTTCAAG 1140  
Db 360 LeuAlaGlnAlaMetSerGlnThrAsnSerThrIleLeuMetGlnArgSerAsnPheLys 379  
QY 1141 GGCCTCCGCGGATCGTCAAGTCTTCACTGCGGAGAGAGGCGCACTCCCGCGCAAC 1200  
Db 380 GlyProLysArgIleValLysCysPheAsnGlySerGlyGlnGlyHisIleAlaLysAsn 399  
QY 1201 TGCCTGGCGCGCGCGAGAGGCTGCTGGAAGTGGCGAGAGGCGCGCAAGTGAAG 1260  
Db 400 CysArgAlaProArgGlySerGlyCysTrpLysCysGlyLysGlnGlyHisGlnMetLys 419  
QY 1261 GACTGACCGGAGCGCGAGCGCAACTCTGGGCGAAGATGCGCGAGCGAGCGCGC 1320  
Db 420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439  
QY 1321 CCGGCGCACTTCTGCGAGCGCGCGCGAGCGCGAGCGCGCGCGCGAGACTTCCGC 1380  
Db 440 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGlnSerPheArg 459  
QY 1381 TTCGAGAGACCAACCCCGCGCGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Db 460 PheGlnGluThrThrProAlaLeuGlnGlnGlyProLysAspArgGluProLeuThrSer 479  
QY 1441 CTGAAGAGCTGTGGCGAGAGAGCGCGCGTGAAGCAG 1476  
Db 480 LeuArgSerLeuPheGlySerAspProLeuSerGln 491

RESULT 8  
US-10-339-217-104

; Sequence 104, Application US/10339217  
; Publication No. US20030198648A1  
; GENERAL INFORMATION:  
; APPLICANT: Buechter, Douglas  
; APPLICANT: Hou, Xiaohong

; APPLICANT: Maior, Christopher W.  
; APPLICANT: Rice, William G.  
; APPLICANT: Yang, Wengang  
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of  
; TITLE OF INVENTION: Nucleospsid 7 Protein to HIV-1 RNA  
; FILE REFERENCE: 111021.143 (ACH-US1)  
; CURRENT APPLICATION NUMBER: US/10/339,217  
; CURRENT FILING DATE: 2003-01-09  
; PRIOR APPLICATION NUMBER: US 60/347,369  
; PRIOR FILING DATE: 2002-01-11  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 104  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: HIV-1  
; US-10-339-217-104

Alignment Scores:  
Pred. No.: 7,45e-119 Length: 491  
Score: 2407.50 Matches: 449  
Percent Similarity: 96.34% Conservative: 25  
Best Local Similarity: 91.26% Mismatches: 17  
Query Match: 86.85% Indels: 1  
DB: 14 Gaps: 1

US-09-475-704a-3 (1-1479) x US-10-339-217-104 (1-491)

QY 1 ATGGCGCGCGCGCGCGAGCATCTGCGCGCGCGAGCGAGTGAAGCGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCGCGCGCGCGAGAGTGTCAATGATGAAGCACTGTGTGTGGCGAGCGCGAG 120  
Db 21 LeuArgProGlyGlyArgGlySerIleTyrMetLeuLysHisLeuValTrpAlaSerArgGlu 40  
QY 121 CTGGAGAGTTCGGCGCTGAACCCCGCGCTGTGAAGCAAGCGCGCGCTGCAAGCATC 180  
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnGlyCysLysGlnIle 60  
QY 181 ATCCGCGAGCTGCAACCCCGCGCTGCGAGCGCGAGCGAGAGTGAAGCGCTTCAAC 240  
Db 61 IleLysGlnLeuHisProAlaLeuLysThrGlyThrGlnGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGGCGCGCGCTGTACTGCTGCGAGAGATCGAGGTCGCGCGACACCGAGAGGCC 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnAsnIleGlnValArgAspThrLysGlnAla 100  
QY 301 CTGGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGlnGluGlnGlnAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCCGAGAGGCGAGAGTGAAGCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 420  
Db 121 AlaAspGlnGly--ValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyMet 139  
QY 421 GTGCAACGAGCGATCAGCGCGCGCGAGCGCGTGAAGCGCTGGTGAAGTGAAGAGAG 480  
Db 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLys 159  
QY 481 GCCTTCAAGCGCGAGAGTATCCCATGTCACCGCGCTGAGCGAGCGCGAGCGCGAG 540  
Db 160 AlaPheSerProGluValIlePrometPheThrAlaLeuSerGluGlyAlaThrProGln 179  
QY 541 GACCTGAACGATGTTGAACACCGTGGCGCGCGAGCGCGAGCGCGCATGCGATGCTGAAG 600  
Db 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLys 199  
QY 601 GACACCATCAACAGAGAGCGCGCGAGTGGAGCGCGGTGACACCCCGTGAACCGCGCGCC 660  
Db 200 AspThrIleAsnGlnGlnAlaIleGlnTrpAspArgLeuHisProValHisAlaGlyPro 219  
QY 661 ATCGCCCGCGCGAGTGGCGAGCGCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGC 720

```

Db      220 AlaAlaProGlyGlnMetArgIuProArgGlySerAspIleAlaGlyThrHisSerThr 239
Qy      721 CTGCAGAGACGATCCCTGATGACCAAGACACCCCACTCCCGTGGGCGACATTCAC 780
Db      240 LeuGlnGluGlnIleAlaThrMetThrGlyAsnProProValGlyAspIleTyr 259
Qy      781 AAGCGGTGATCATCTGGGCGCGAACAAGATGTCGGATGACAGCCCGGAGCATC 840
Db      260 LysArgTrpIleIleuGlyLeuAsnLysIleValArgMetCysSerProValSerIle 279
Qy      841 CTGCATCATACAGAGGCGCCCAAGAGCCCTTCCGACTACGTGAGCCGCTTCTTCAG 900
Db      280 LeuAspIleIleGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299
Qy      901 ACCCTGGCGCGGAGAGACACCCAGAGGTGAAGATGTGATGACCAACCTCTGCTG 960
Db      300 ValLeuArgAlaGlnGlnIleAlaThrGlnAspValLysAsnTrpMetThrAspThrLeu 319
Qy      961 GTGCAGAACGCGAACCCCGACTGCAGACATCTCTGGGCGCTGGCGCGCGCGCGCAGC 1020
Db      320 IleGlnAsnAlaAsnProAspCysLysTrpIleLeuLysAlaLeuGlyProAlaAlaSer 339
Qy      1021 CTGCAGAGATGATGACCCCTGCGCAGGCGTGGCGCGCGCGCGCAAGCGCGCGCTG 1080
Db      340 LeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgVal 359
Qy      1081 CTGGCCGAGGACATGAGCCAGCGCAACACACAGCGTGATATGACGAAGACCACTTCAG 1140
Db      360 LeuAlaGlnLysMetSerGlnIleAsnSerAsnIleMetMetGlnArgSerAsnPheLys 379
Qy      1141 GGGCGCGCGGCGATGTCAGAGTCTCAACTGCGGCAAGAGGCGCACTCCCGCGCAC 1200
Db      380 GlySerLysArgLysIleValLysCysPheAsnCysGlyLysGlnGlyHisIleAlaArgAsn 399
Qy      1201 TGGCGCGCGCGCGCAAGAGGCTGCTGGAAGTGGCGGAGGCGGCGCAAGATGAAG 1260
Db      400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGlnGlyHisGlnMetLys 419
Qy      1261 GACTGACCGGACCGCGCAAGGCGCACTTCTGGGCAAGATGTGGCGCAAGCGCGCGCGC 1320
Db      420 AspCysThrGlnArgGlnIleAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439
Qy      1321 CCGCGCACTTCTCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCTTCGCGC 1380
Db      440 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGlnSerPheArg 459
Qy      1381 TTGCAGAGACCAACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db      460 PheGlnGluThrThrProAlaProLysGlnGlnLysSerLysAspArgGluProLeuIleSer 479
Qy      1441 CTGAAGAGCTGTGGCAACGACCCCTGAGCGCAG 1476
Db      480 LeuLysSerLeuPheGlySerAspProSerSerGln 491

```

RESULT 9  
US-10-339-217-143

```

; Sequence 143, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Maijor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT APPLICATION NUMBER: US/10/339,217
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 144

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 143
; LENGTH: 491
; TYPE: PRT
; ORGANISM: HIV-1
; US-10-339-217-143

Alignment Scores:
pred. No.:      2,83e-118      Length:      491
Score:          2396.50      Matches:      449
Percent Similarity: 95.53%      Conservative: 21
Best Local Similarity: 91.26%      Mismatches: 21
Query Match:      86.45%      Indels:      1
DB:              14          Gaps:      1

US-09-475-704A-3 (1-1479) x US-10-339-217-143 (1-491)
Qy      1 ATGGGCGCGCGCGCGAGCATCTCTGGCGCGGCAAGTGAACGCTGGAGCGCATCCGC 60
Db      1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTrpLysIleArg 20
Qy      61 CTGGCGCGCGCGCGCAAGAGTCTATGATGAAGACCTGTGTGGCGCGCGCGAG 120
Db      21 LeuArgProGlyGlyLysLysIleTyrMetIleLysIleValTrpAlaSerArgGlu 40
Qy      121 CTGCAGAGTTCGCGCTGAACCCCGCGCTCTGAGACCGGACGAGAGCTGCAAGCGATC 180
Db      41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerGlnGlyCysLysGlnIle 60
Qy      181 ATCCGCGAGTTCACACCCGCGCTGCAAGACCGGACGAGAGAGCTGAAGAGCTGTTCAC 240
Db      61 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLysLeuArgSerIleHisAsn 80
Qy      241 ACCGTGGCGCACCTGTACTGCTGTGACAGAAAGATGAGTTCGCGACACCAAGAGGCC 300
Db      81 ThrValAlaThrLeuTyrCysValHisAlaGlyIleGluIleArgAspThrLysGlnAla 100
Qy      301 CTGCAGAGTGAAGAGAGAGAGAGAAAGATGCCAGAGAAATCAGAGGCGCGAGGCC 360
Db      101 LeuAspLysIleGlnGlnGlnGlnLysSerGlnGlnLysThrGlnIleAlaLysGlu 120
Qy      361 GCGCAAGAGGCGCAAGTGAAGCAGAACTACCCCATGTGTCAGAACCTGAGGCGCAGATG 420
Db      121 AlaAsp---GlyLysValSerGlnAsnTrpIleValGlnAsnLeuGlnGlnMet 139
Qy      421 GTGCACAGAGCGCATAGCCCGCGCACCTGAACGCGCTGGGTGAAGTGAAGAGAGAG 480
Db      140 ValHisGlnAlaLeuSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLys 159
Qy      481 GCGTTCAGCGCGGAGATGATCCCAATGTTACCCGCGCTGAGGAGGCGCGCACCCCGCAG 540
Db      160 AlaPheSerProGlnIleIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGln 179
Qy      541 GACTGGAACAGATGTGAACACGTTGGGCGGCGCACAGCGCGCATGAGATGCTGAAG 600
Db      180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199
Qy      601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCGCGTGCACCCGCTGACCGCGCGCGCC 660
Db      200 AspThrIleLeuGlnGlnLysAlaAlaGlnTrpArgArgLysHisLysProAlaGlnIleArg 219
Qy      661 ATCCGCGCGCGCAATGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAGC 720
Db      220 IleAlaProGlyGlnMetArgGluProArgLysSerAspIleAlaGlyThrThrSerThr 239
Qy      721 CTGCAGAGAGATGCTGTGATGACAGAACCCCGCATCCCGTGGGCGCGACATTCAC 780
Db      240 LeuGlnGlnGlnIleAlaTrpMetThrGlyAsnProProValGlyGlyLysIleTyr 259
Qy      781 AAGCGGTGATCATCTGGGCGCTGAACAAGATGTGCGAGTATACGCGCGCGTGAAGCATC 840
Db      260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetCysSerProValSerIle 279

```







Db 381 GlyProLysArgThrValLysCysPheAnaCysGlyLysGluGlyHisIleLeuAlaArgSer 400

QY 1201 TGCCGCGCCCCCGCGCAAGAGGGCTGCTGGAAGTGGCGGCAAGAGGGCCACCAATGAAG 1260

Db 401 CysArgIleProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleIleMetLys 420

QY 1261 GACTGCACCGAGCGCCAGGCGCACTTCGTGGCGCAATGTGGCCAGGACACAGAGGGCCGC 1320

Db 421 AspArgThrGluArgGlnIleAsnPheLeuGlyLysIleTrpProSerTyrLysGlyArg 440

QY 1321 CCGGCGAACTTCCTGCGAGAGCGGCGCCGAGGCCACCGCCCCCGCCGAGAGCTTCGGC 1380

Db 441 SerIleAsnMetLeuGlnSerIleArgProGluProSerAlaProProIleGluSerPheArg 460

QY 1381 TTCGAGAGAGCAACCCCGCGCCAGAGAGCGAGAGCAGAGACCGCGAG---ACCTGAGC 1437

Db 461 PheGluGluLys-----ArgGluProLysAspLysGluProProLeuThr 474

QY 1438 AGCCTGAAGAGCCTGTTCCGCAACGACCCCTGAGGCCAG 1476

Db 475 SerLeuLysSerLeuPheGlySerAlaProSerSerGln 487

```

RESULT 12
US-10-339-217-107
; Sequence 107, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds which Inhibit Binding of
; FILE REFERENCE: 111021.143(ACH-US1)
; CURRENT FILING DATE: US/10/339,217
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: US 60/347,369
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107
; LENGTH: 497
; TYPE: PRT
; ORGANISM: HIV-1
; US-10-339-217-107

```

Alignment Scores:	
Pred. No.:	1,74e-117
Score:	2381.50
Percent Similarity:	94.79%
Best Local Similarity:	90.18%
Query Match:	85.91%
DB:	14
Length:	497
Matches:	450
Conservative:	23
Mismatches:	17
Indels:	9
Gaps:	3

US-09-475-704A-3 (1-1479) X US-10-339-217-107 (1-497)

QY	1	ATGGGGGCGCCGGCCAGCATCTCTGGCGGGCGCAAGCTGAGCGCCTGGAGCGCATCCG	60	
		1	MelGlyAlaIArgAlaSerIleLeuSerGlyGlyIysIleAspLysITPGLuArgIleArg	20
Db				
QY	61	CTGGCGCCCGCGCGCGCAAGAGTCAACATGATGAAGACGACTGGTGGCGACCGCGAG	120	
		21	LeuArgProGlyGlyIysIysHisIlyrMetLeuLysHisIleValITPAlaSerArgIu	40
Db				
QY	121	CTGGAGAGTTCCGCCCTGAAACCCCGGCGCTGTGAGACGACGAGGCGTGCAGACATC	180	
		41	LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluITrAlaGluGlyCysIysGlnIle	60
Db				
QY	181	ATCCGCGACCTGCACCCCGCCTCTGCAGACCGGACCGAGAGCTGAAAGCGCTGTTCAAC	240	
		61	IleLysGlnLeuGlnProAlaLeuGlnIlnhGlyIThrGluGlnLeuArgSerLeuPheAsn	80

QY	241	ACCGTGGCCACCTCTGTACTGGCTGTGACAGAAAGATCGAGTCTCCGACACCAAGAGGCTC	300
Db	81	ThrValAlaThrLeuTryCysValHisValysGlyIleGluValAlaArgAspThrTyrSGluAla	100
QY	301	CTGGACAAGATGAGAGAGAGAGACAAGATGCTCCAGCAAGATGATCAGAGGCTCAGGCTC	360
Db	101	LeuAspArgTyrIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu	120
QY	361	GCCGACAAAGGCGCAAGTGAAGCCAGAACTAACCCATCTGGCAGAACCTTGAGGGCCAGATG	420
Db	121	AlaAspArgGlu--LysValSerGlnAsnTyrProIleValGlnAsnAlaGlnIleGlnMet	139
QY	421	GTGCACACGAGCCATCAGGCCCCGACGCTCGAAACGCTGGGTGAAGTATGATCAGAGAGAG	480
Db	140	ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluTyr	159
QY	481	GCTTCAGACCCCGAGGTATCCCATGTTCAACGCTGAGCGAGGAGGCGCAACCCCGCAG	540
Db	160	AlaPheAsnProGluValIleProMetPheThrAlaLeuSerGluGluYAlaThrProGln	179
QY	541	GACCTGAACAAGATGTGTAACAACGTGGGCGGCGCAACGCGCCCATCAGATGTGTGAG	600
Db	180	AspLeuAsnThrMetLeuAsnThrValJalGlyHisGlnAlaAlaMetGlnMetLeuLys	199
QY	601	GACACCATCAACGAGAGGCGCGGAGTGGAGCCGCTGACACCCGCTCAGACGCGCGGCCCC	660
Db	200	AspThrIleAsnGluGluAlaIleGluTrpAspArgThrHisProValHisAlaGluTyr	219
QY	661	ATCGGCCCCCGGCAGATTCGCGAGGCCCCGCGGCGACGACATCGCGCGGCAACCAACGAC	720
Db	220	ValAlaProGlyGlnMetArgGluProArgIleSerAspIleAlaGluTyrThrSerThr	239
QY	721	CTGCAGAGCAGATGCGCTGTGATACACGCAACCCCGCATCCCGTGGGAGCATCTAC	780
Db	240	LeuGlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleTyr	259
QY	781	AAGCGGTGATCATCTTGGGCTTGAACAAGATCGTGGAGTGAACGCCCCGTGAGCATC	840
Db	260	LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProValSerIle	279
QY	841	CTGCACATCAAGACGAGGCGCCCAAGAGGCGCTTCGCGGCACTACGAGGACCGTCTTTCAG	900
Db	280	LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys	299
QY	901	ACCTGTGCGCGCGACAGACGACCCAGAGAGGTGAAGAACTGATGATCCGACACCTGCTG	960
Db	300	ThrLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu	319
QY	961	GTGCAGAACCGCAACCCCGCACTGGAAGCACTCTGGCGGTCTTCGCGCGCGGCGCGAC	1020
Db	320	ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGluYAlaSer	339
QY	1021	CTGGAGGAGATGATGACCGCTGTGACAGGCGCTGGGCGGCGCCACGACCAAGAGCCGCGCTG	1080
Db	340	LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysValAlaArgAla	359
QY	1081	CTGGCCGAGGCGATGAGCGCAGGCGCAACACAGCGTATGATGCAAGAGCACTTCAAG	1140
Db	360	LeuAlaGluAlaMetSerGlnThrAsnSerAsnIleLeuValGlnArgSerAsnPheLys	379
QY	1141	GAGCCCGGCGCATCGTCAAGTGTCTTCAACCTGCGGCAAGAGAGGCGCATGATGCGCGCAC	1200
Db	380	GlySerAsnArgIleValLysCysPheAsnCysGluLysValGlyHisIleValAlaArgAsn	399
QY	1201	TGCGGCGCCCCCGCAAGAGGCTGTGTGAAGTGCAGCAAGAGGCGCAACAGATGAG	1260
Db	400	CysArgAlaProArgLysGlyGlyCysTrpLysCysGlyGlnGluGlyHisGlnMetLys	419
QY	1261	GACATGACCGAGCGCGCGGCGCAATTCCTGGGCAAGATCTGGCCGACCGCAAGGCGCGC	1320
Db	420	AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg	439
QY	1321	CCCGGCAACTTCTGCAAGCGCGCCC-----GAGCCCAACGCGC	1359



```
|||||
Db      440  ProGlyAsnPhenLeuInAsnArgProGluProThrAlaProProAlaGluProThrAla 459
QY      1360  CCCCCCGCGAGAGCTTCCGCTTCGAGAGACCAACCCCGCGCGAAGCAAGAGACAG 1419
Db      460  ProProAlaGluSerPheArgPheGluGluThrThrProAlaProLysArgGlu---Lys 478
QY      1420  GACCGCGAGACCTGACGACCTGAGAGAGCTGTGCGGCAAGACCCCGTGAAGCCAG 1476
Db      479  GluArgGluProLeuThrSerLeuLysSerLeuPheGlyAsnAspProSerSerGln 497

RESULT 13
US-10-339-217-105
; Sequence 105, Application US/10339217
; Publication No. US20030198648A1
; GENERAL INFORMATION:
; APPLICANT: Buechter, Douglas
; APPLICANT: Hou, Xiaohong
; APPLICANT: Marlor, Christopher W.
; APPLICANT: Rice, William G.
; APPLICANT: Yang, Wengang
; TITLE OF INVENTION: Methods for Identifying Compounds Which Inhibit Binding of
; FILE REFERENCE: Nucleocapsid 7 Protein to HIV-1 RNA
; CURRENT FILING DATE: 2003-01-09
; PRIOR FILING DATE: 2002-01-11
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105
; LENGTH: 508
; TYPE: PRT
; ORGANISM: HIV-1
US-10-339-217-105

Alignment Scores:
Pred. No.: 1,64e-115 Length: 508
Score: 2344.00 Matches: 446
Percent Similarity: 93.328 Conservative: 29
Best Local Similarity: 87.624 Mismatches: 16
Query Match: 84.564 Indels: 18
DB: 14 Gaps: 4

US-09-475-704a-3 (1-1479) x US-10-339-217-105 (1-508)
QY      1  ATGGGCGCGCGCGCGCATCTGCGCGGCGGCAAGCTGACGCTGAGAGCGATCCG 60
Db      1  MetGlyAlaArgAlaSerIleLeuArgGlyThrLysLeuAspAlaTrpGluLysIleArg 20
QY      61  CTGGCGCGCGCGCGCGCAAGAGCTACATGATGAAGACACCTGTGTGGCGCGCGCGAG 120
Db      21  LeuArgProGlyGlyLysLysValIleTyrMetLeuLysValIleValIleTrpAlaSerArgGlu 40
QY      121  CTGAGAGAGTTGCGCTGAACCCCGCGCTGTGAGAGACCGAGGCGGTGCAAGAGATC 180
Db      41  LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluLysCysLysGlnIle 60
QY      181  ATCCGCGACCTGACACCCCGCGCTGACAGCCGCGAGAGAGAGCTGAGAGAGCTGTTCAC 240
Db      61  MetLysGlnLeuHisProAlaLeuGlnThrGlyThrGluLysSerLeuTyrAsn 80
QY      241  ACCGTCGCGACCTGTACTGCTGTGCAAGAAAGATGAGAGTCCGCGACCAAGAGGCC 300
Db      81  ThrValAlaThrLeuLysCysValHisGluAsnIleLysValArgAspThrLysGluAla 100
QY      301  CTGGAACAATCGAGAGAGAGCAAGAAC-----AAGTCGAGAGAGAAATCCAGAGGCC 354
Db      101  LeuAspLysIleGluLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 120
QY      355  GAGGCGCGCGAGAGAGAGAGTGAAGCAATCAATCCATCGTGTGAGAGAGGCC 414
Db      121  LysValAlaAspGlu---LysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGly 139
```

```
QY      415  CAGATGTCACAGGCGCATGACCCCGCGACCCCTGAACGCTGGTGAAGATGAGAG 474
Db      140  GlnMetValHisGlnAsnMetSerProArgThrLeuAsnAlaTrpAlaLysValIleGlu 159
QY      475  GAGAAAGCTTCAGCCCGAGGTGATCCCATGTTCAACCGCGTGAAGAGAGGCCACCC 534
Db      160  GluLysAlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThr 179
QY      535  CCGCAGACCTGAACACGATGTTGAACACCGTGGCGCGCCACAGGCCCGCATGCAATG 594
Db      180  ProGlnAspLeuSerThrMetLeuAsnThrValGlyGlnIleAlaIleMetGlnMet 199
QY      595  CTGAAGGACACATCAACAGAGAGGCGCGAGTGGGACCGGTCGACCCCGCGAGGCC 654
Db      200  LeuLysAspThrIleAsnGluLysAlaAlaGluTrpAspArgLeuHisProValHisAla 219
QY      655  GCGCCCATGCGCCCGCGAGATGCGCGAGACCCCGCGAGAGACATCCCGCGACACCC 714
Db      220  GlyProMetAlaProGlyGlnLeuArgGluProArgGlySerAspIleAlaGlyThrThr 239
QY      715  AGCACCCTGCAAGAGACGATCGCTGATGACCAAGACCCCGCATCCCGTGGCGAC 774
Db      240  SerThrLeuArgGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAsp 259
QY      775  ATCTCAAGCGGTGATCATCTGGGCGTGAACAAGATGTGCGGATGACAGCCCGGTG 834
Db      260  IleTyrLysArgTrpIleLeuGlyLeuAsnLysIleValArgMetLysSerProVal 279
QY      835  AGCATCTGAGATCAAGAGAGGCGCCCAAGAGACCTTCGCGACATCACTGAGCCGCTTC 894
Db      280  SerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspLysValAspArgPhe 299
QY      895  TTCAGACCTGCGCGCGCGAGACACCCAGAGAGTGAAGAATGTGATGACCGACACC 954
Db      300  PheLysAlaLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrGluThr 319
QY      955  CTGCTGTGCAAGAGCGCAACCCCGCATGACCATCTGCGGCTTCGCGCGCGCGCG 1014
Db      320  LeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyIleGly 339
QY      1015  GCCAGCTGAGAGATGATGACCGCTGCGAGGCGGTGAGCGGCCCGCCAGCAAGGCC 1074
Db      340  AlaThrLeuGluGlnMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysAla 359
QY      1075  CGCGTGTGCGCGAGCGCATAGCCAGCGCC---AACACCGAGGTGATGAGAAAGCC 1131
Db      360  ArgValIleuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetMetGlnArgSer 379
QY      1132  AACTTCAAGGCGCGCGCGCATCGTCAAGTCTTCACTGGGCGCAAGAGGCCCATC 1191
Db      380  AsnPheLysSerSerLysArgIleValLysCysSerAsnCysGlyLysGluLysIle 399
QY      1192  GCCCGCACTGCGCGCGCGCGCGCAAGAGAGGCTCTGGAATGTGCGGCAAGAGGCCAC 1251
Db      400  AlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluLysHis 419
QY      1252  CAGATGAAGACTGCAACGAGCGCGCAAGCGCAACTTCTGGGCAAGATCTGCGCCAGCCAC 1311
Db      420  GlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHis 439
QY      1312  AAGGCGCGCGCGCGCACTTCTGCGAG----- 1338
Db      440  LysGlyArgProGlyAsnPhenLeuInAsnArgProGluProThrAlaProProAlaGlu 459
QY      1339  -----AGCGCGCGCGAGCGCGCGCGCGCGCGAGAGCTTCGCTTGCAGAGAG 1389
Db      460  SerPheAlaArgAsnArgProGluProThrAlaProProAlaGluSerPheArgPheGluGlu 479
QY      1390  ACCACCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1449
Db      480  ThrThrProThrProLysGlnGluProLysAspArgAspProLeuThrSerLeuLysSer 499
```





**THIS PAGE BLANK (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:05:25 ; Search time 41.8258 Seconds  
(without alignments)  
5279.323 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 2772  
Sequence: 1 atggcgccgcgcgcagcat.....acgaccctcgagcagtaa 1479

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=x1p  
-Q=/cg2\_1/USPTO.spool.p/US09475704/runtac.10032005.140222.14851/app\_query.fasta\_1.3342  
-DB=Issued\_Patents\_AA -QFMT=fastaen -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORExt -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704 @CGN 1\_1.86 @runtac.10032005.140222.14851 -ICPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELDP=6 -DELEXT=7

Database :

Issued\_Patents\_AA:\*  
1: /cg2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cg2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cg2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cg2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cg2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cg2\_6/prodata/1/1aa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2435	87.8	492	4	US-09-991-258-5
2	2256	81.4	500	2	US-08-816-1558-45
3	2256	81.4	500	3	US-09-079-587-45
4	2249	81.1	500	1	US-08-375-510-1
5	2249	81.1	500	2	US-08-487-657-1
6	2249	81.1	500	4	US-09-309-572-16
7	2249	81.1	500	4	US-09-718-096-16
8	2245	81.0	1350	4	US-09-952-060-35
9	2235	80.6	512	3	US-08-463-210-8
10	2235	80.6	512	3	US-09-124-900-2
11	2235	80.6	512	4	US-08-463-028-8
12	2106	76.0	493	4	US-09-952-060-33

13	2037.5	73.5	478	1	US-08-127-499A-11	Sequence 11, Appl
14	2037.5	73.5	478	1	US-08-482-847-11	Sequence 11, Appl
15	2023	73.0	437	3	US-08-392-794A-2	Sequence 2, Appl1
16	1976	71.3	512	4	US-09-319-588C-4	Sequence 4, Appl1
17	1797.5	64.8	498	1	US-08-470-202-59	Sequence 59, Appl1
18	1797.5	64.8	498	1	US-08-471-770-59	Sequence 59, Appl1
19	1797.5	64.8	498	2	US-08-468-059-59	Sequence 59, Appl1
20	1797.5	64.8	498	3	US-09-109-916-59	Sequence 59, Appl1
21	1797.5	64.8	498	4	US-09-886-156-59	Sequence 59, Appl1
22	1797.5	64.8	498	4	US-09-886-149-59	Sequence 59, Appl1
23	1797.5	64.8	498	4	US-09-886-150-59	Sequence 59, Appl1
24	1797.5	64.8	498	4	US-09-886-159-59	Sequence 59, Appl1
25	1797.5	64.8	498	4	US-10-326-090-59	Sequence 59, Appl1
26	1775.5	64.1	498	1	US-08-470-202-60	Sequence 60, Appl1
27	1775.5	64.1	498	1	US-08-471-770-60	Sequence 60, Appl1
28	1775.5	64.1	498	2	US-08-468-059-60	Sequence 60, Appl1
29	1775.5	64.1	498	3	US-09-109-916-60	Sequence 60, Appl1
30	1775.5	64.1	498	4	US-09-886-156-60	Sequence 60, Appl1
31	1775.5	64.1	498	4	US-09-886-149-60	Sequence 60, Appl1
32	1775.5	64.1	498	4	US-09-886-150-60	Sequence 60, Appl1
33	1775.5	64.1	498	4	US-09-886-159-60	Sequence 60, Appl1
34	1775.5	64.1	498	4	US-10-326-090-60	Sequence 60, Appl1
35	1681	60.6	363	2	US-08-050-478-130	Sequence 130, App
36	1681	60.6	363	2	US-08-850-049-130	Sequence 130, App
37	1681	60.6	363	3	US-09-414-117-130	Sequence 130, App
38	1681	60.6	363	4	US-09-678-437-130	Sequence 130, App
39	1681	60.6	363	4	US-09-943-722-130	Sequence 130, App
40	1536.5	55.4	458	1	US-07-648-796A-5	Sequence 5, Appl1
41	1536.5	55.4	739	1	US-07-648-796A-7	Sequence 7, Appl1
42	1530.5	55.2	337	1	US-07-648-796A-1	Sequence 1, Appl1
43	1514.5	54.6	328	4	US-08-776-188C-77	Sequence 77, Appl1
44	1478.5	53.3	518	4	US-09-206-551-45	Sequence 45, Appl1
45	1413	51.0	294	3	US-09-370-368-3	Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-991-258-5  
; Sequence 5, Application US/0991258  
; Patent No. 6783939  
; GENERAL INFORMATION:  
; APPLICANT: Olmsted, Robert  
; APPLICANT: Keith, Paula  
; APPLICANT: Dryga, Sergey  
; APPLICANT: Caley, Ian  
; APPLICANT: Maughan, Maureen  
; APPLICANT: Johnston, Robert  
; APPLICANT: Davis, Nancy  
; APPLICANT: Swastrom, Ronald  
; TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE  
; FILE REFERENCE: 0113.0001U3  
; CURRENT APPLICATION NUMBER: US/09/991,258  
; PRIOR FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: 09/902,537  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: 60/216,995  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; No. 6783939 =  
; US-09-991-258-5  
; OTHER INFORMATION: synthetic construct  
Alignment Scores: 2,97e-141 Length: 492  
Pred. No.: 2435.00 Matches: 458  
Score:

Percent Similarity: 96.96% Conservative: 20  
Best Local Similarity: 92.90% Mismatches: 13  
Query Match: 87.84% Indels: 2  
DB: 4 Gaps: 2  
US-09-475-704A-3 (1-1479) x US-09-991-258-5 (1-492)

QY 1 ATGGGGGGGGGGGGGGGATCTTGGGGGGGGGCAAGCTGAGCCCTGGAGGCGATCCGC 60  
DB 1 MetAlaAlaArgAlaSerIleLeuArgGlyGluLeuAspArgTrpGluYsrIleArg 20  
QY 61 CTGGCGCCCGGGCGGCAAGAGTGTCTACATGATGAAGCACTGGTGTGGCGAGCCGCGAG 120  
DB 21 LeuArgProGlyGlyYsrYsrIleYsrMetLeuYsrIleValTrpAlaSerArgGlu 40  
QY 121 CTGGAGAGTTCGCTGAAAGCCCGGCTGTGGAGACCAAGGAGGCTGGACAGATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGluYsrGlySerIle 60  
QY 181 ATCCGCGACCTGACCCCGCCCTGGACAGCCGAGAGGAGGAGTGAAGGCTGTTCAC 240  
DB 61 MetYsrGlnLeuGlnProAlaLeuGlnTrpGlyThrGluGluYsrLeuYsrLeuYsr 80  
QY 241 ACCGTGGCCACCTGTACTGCTGACAGAGAGATGAGGTCCGCGACCAAGAGGCGC 300  
DB 81 ThrValAlaThrLeuYsrYsrValAlaSerGluYsrIleGluValArgAspThrYsrGlu 100  
QY 301 CTGGACAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 101 LeuAspYsrIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120  
QY 361 GGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
DB 121 AlaAsp--GlyYsrValSerGlnAsnYsrProIleValGlnAsnLeuGlnGlnMet 139  
QY 421 GTGCAACAGGCGATGAGCCCGGCGACCTGAAAGCTGGGTGAGTGTGAGAGAGAG 480  
DB 140 ValIleGlnAlaIleSerProArgThrLeuAsnAlaTrpValYsrAlaIleGluGluYsr 159  
QY 481 GCGTTGAGCCCGGAGGTGATCCCATGTTTCAAGCCCTGAGAGGAGGAGGAGGAGG 540  
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluYsrIleAlaThrProGln 179  
QY 541 GACCTGAACAGAGTGTGAACACCGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 600  
DB 180 AspLeuAsnTrpMetLeuAsnTrpValGlyYsrGlnAlaAlaMetGlnMetLeuYsr 199  
QY 601 GACACCATCAACAGAGAGGCGCGGAGTGGAGCCGCGTGCACCCCGTGCAGCCGCGCC 660  
DB 200 AspTrpIleAsnGluGluAlaIleGluTrpAspArgLeuAlaProValAlaIleGlyPro 219  
QY 661 ATCCGCGCCCGGCGAGTGGCGAGCCCGGCGGAGGAGATGCGCGGACCAAGCAGC 720  
DB 220 IleAlaProGlyGlnMetArgGlnProAspArgIleAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGGAGAGAGAGATGCGCTGTGATGACAGCAAGCCCGCATCCCGTGGGAGCATTC 780  
DB 240 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProIleProValGlyAspIleYsr 259  
QY 781 AAGCGGTGATCATCTGGGCTGAAAGAGATGTCGCGATGACAGCCCGCTGAGCATC 840  
DB 260 YsrArgTrpIleIleLeuGlyLeuAsnYsrIleValArgMetYsrSerProValSerIle 279  
QY 841 CTGGACATCAACAGAGGCGCGCAAGAGCCCTTCGCGATTAAGTGAAGCGCTTTCAG 900  
DB 280 LeuAspIleArgGlnGlyProYsrGlnProPheArgAspIleValAspArgPhePheYsr 299  
QY 901 ACCGTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 300 ThrLeuArgAlaGlnIleAlaThrGlnGluValYsrAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGCAAGAGGCAAGCCCGATGCAAGACATCTGCGGCTCTCGGCGCGCGCGCAG 1020  
DB 1021 CTGGAGAGATGATGACCGCTGCGAGGCGGTGGCGGCGCGCGCGCGCGCGCGCG 1080  
DB 340 LeuGluGlnMetThrAlaCysGlnGlyValGlyYsrProGlyYsrIleYsrAlaArgVal 359  
QY 1081 CTGGCGGAGGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1137  
DB 360 LeuAlaGlnAlaMetSerGlnThrAsnSerGlyAsnIleMetGlnArgSerAsnPhe 379  
QY 1138 AAGGCGCCCGGCGATGCTCAAGTGTCTTCACTGCGGCAAGAGAGGAGGAGGAGGAG 1197  
DB 380 YsrGlyProArgArgIleValYsrCysPheAsnYsrGlyYsrGluGlyYsrIleAlaArg 399  
QY 1198 AACTCGCGCGCCCGCGCAAGAGGCGTGTGAAAGTGGCGGAGAGGAGGAGGAGGAG 1257  
DB 400 AsnYsrAlaProArgYsrYsrYsrYsrYsrYsrYsrYsrYsrYsrYsrYsrYsrYsr 419  
QY 1258 AAGACATGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1317  
DB 420 YsrAspYsrThrGluArgGlnAlaAsnPheLeuGlyYsrIleTrpProSerThrYsrGly 439  
QY 1318 CGCCCGGCAACTTCTCTGAGAGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1377  
DB 440 ArgProGlyAsnPheLeuGlnAsnArgProGluProTrpAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
DB 460 ArgPheGluGluThrThrProAlaProYsrGlnGluProIleGluArgGluProLeuThr 479  
QY 1438 AGCGTGAAGAGGCTGTGGGCAAGAGCCCGTGGAGCGCAG 1476  
DB 480 SerLeuYsrSerLeuPheGlySerAspProLeuSerGln 492

DB 320 ValGlnAsnAlaAsnProAspCysYsrThrIleLeuAlaGlnLeuGlyProGlyAlaThr 339  
QY 1021 CTGGAGAGATGATGACCGCTGCGAGGCGGTGGCGGCGCGCGCGCGCGCGCGCG 1080  
DB 340 LeuGluGlnMetThrAlaCysGlnGlyValGlyYsrProGlyYsrIleYsrAlaArgVal 359  
QY 1081 CTGGCGGAGGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1137  
DB 360 LeuAlaGlnAlaMetSerGlnThrAsnSerGlyAsnIleMetGlnArgSerAsnPhe 379  
QY 1138 AAGGCGCCCGGCGATGCTCAAGTGTCTTCACTGCGGCAAGAGAGGAGGAGGAGGAG 1197  
DB 380 YsrGlyProArgArgIleValYsrCysPheAsnYsrGlyYsrGluGlyYsrIleAlaArg 399  
QY 1198 AACTCGCGCGCCCGCGCAAGAGGCGTGTGAAAGTGGCGGAGAGGAGGAGGAGGAG 1257  
DB 400 AsnYsrAlaProArgYsrYsrYsrYsrYsrYsrYsrYsrYsrYsrYsrYsrYsrYsr 419  
QY 1258 AAGACATGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1317  
DB 420 YsrAspYsrThrGluArgGlnAlaAsnPheLeuGlyYsrIleTrpProSerThrYsrGly 439  
QY 1318 CGCCCGGCAACTTCTCTGAGAGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 1377  
DB 440 ArgProGlyAsnPheLeuGlnAsnArgProGluProTrpAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
DB 460 ArgPheGluGluThrThrProAlaProYsrGlnGluProIleGluArgGluProLeuThr 479  
QY 1438 AGCGTGAAGAGGCTGTGGGCAAGAGCCCGTGGAGCGCAG 1476  
DB 480 SerLeuYsrSerLeuPheGlySerAspProLeuSerGln 492

RESULT 2  
US-08-816-155B-45  
Sequence 45, Application US/08816155B  
Patent No. 598091  
GENERAL INFORMATION:  
APPLICANT: TARTAGLIA, JAMES  
APPLICANT: COX, WILLIAM I.  
APPLICANT: GETTIG, RUSSELL R.  
APPLICANT: MARTINEZ, HECTOR  
APPLICANT: PAOLETTI, ENZO  
APPLICANT: PINCUS, STEVEN E.  
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
STREET: 745 FIFTH AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10151  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816.155B  
FILING DATE: 12-MAR-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: KOMALSKI, THOMAS J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454310-2990  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:



SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,587  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/816,155  
FILING DATE: 12-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: KOMALSKI, THOMAS J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454310-2990  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ. ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 500 amino acids  
TYPE: amino acid  
STRANDEDNESS: n/a  
TOPOLOGY: linear  
MOLECULE TYPE: amino acid  
US-09-079-587-45

Alignment Scores:  
Pred. No.: 2,59e-130 Length: 500  
Score: 2256.00 Matches: 424  
Percent Similarity: 92.42% Conservative: 39  
Best Local Similarity: 84.63% Mismatches: 28  
Query Match: 81.39% Indels: 10  
DB: 3 Gaps: 5  
US-09-475-704a-3 (1-1479) x US-09-079-587-45 (1-500)

QY 1 ATGGGGGCGCGCGCGCGCATCTGCGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspArgTyrPheValIleArg 20  
QY 61 CTGCGCGCGCGCGCGCAAGAGTGTCTCATGATGAAGACCTGATGGCGCGCGCGCG 120  
DB 21 LeuArgProGlyGlyIleValSerGlyTyrIleValSerGlyIleValIleValIleVal 40  
QY 121 CTGAGAGAGTTCGCGCTGAACCCCGCGCTGCTGAGAGACCGAGGCGCTGACAGATC 180  
DB 41 LeuGlyArgPheAlaValAlaAsnProGlyLeuLeuGlyIleValIleValIleVal 60  
QY 181 ATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
DB 61 LeuGlyIleValIleValIleValIleValIleValIleValIleValIleValIleVal 80  
QY 241 ACCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
DB 81 ThrValAlaThrLeuTyrCysValIleValIleValIleValIleValIleValIleVal 100  
QY 301 CTGAGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 101 LeuAspLysIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIle 119  
QY 361 GCCGCAAGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420  
DB 120 AlaAspThrGlyIleValSerIleValIleValIleValIleValIleValIleValIle 139  
QY 412 GCGCGAGTGTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 471  
DB 140 GlyIleMetValIleGlyIleAlaIleSerProArgThrLeuAsnAlaIleValIleVal 159  
QY 472 GAGGAGAGAGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 531  
DB 160 GluGlyIleValIlePheSerProGlyIleValIleProMetPheSerAlaIleValIle 179  
QY 532 ACCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 591  
DB 180 ThrProGlyIleAspLeuAsnThrMetLeuAsnThrValIleGlyIleGlyIleAlaIleVal 199

QY 592 ATGCTGAAGACACCATCAACGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651  
DB 200 MetLeuLysGlyThrIleAlaAsnGlyIleValIleAlaIleValIleValIleValIle 219  
QY 652 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711  
DB 220 AlaGlyProIleAlaProGlyIleMetArgIleProArgIleValSerAlaIleGlyThr 239  
QY 712 ACCAGACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771  
DB 240 ThrSerThrLeuGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIle 259  
QY 772 GACATCTCAAGCGGTGATCATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 831  
DB 260 GluIleTyrLysArgTyrIleIleLeuGlyIleValIleValIleValIleValIleVal 279  
QY 832 GTGAGCATCTGAGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
DB 280 ThrSerIleLeuAspIleArgIleGlyIleProGlyIleProPheArgPheValAspArg 299  
QY 892 TTCTTCAAGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951  
DB 300 PheTyrLysThrLeuArgAlaGlyIleValIleValIleValIleValIleValIleVal 319  
QY 952 ACCCTGCTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
DB 320 ThrLeuValIleGlyIleValIleValIleValIleValIleValIleValIleValIle 339  
QY 1012 GAGCGCGCGCGCGAGAGATGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1071  
DB 340 AlaAlaThrLeuGlyIleGlyIleMetThrAlaCysGlyIleValIleGlyIleProGly 359  
QY 1072 GCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1125  
DB 360 AlaArgValIleAlaGlyIleAlaMetSerGlyIleValIleValIleValIleValIle 379  
QY 1126 AAGAGCAACTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1185  
DB 380 ArgGlyAsnPheArgAsnGlyIleValIleValIleValIleValIleValIleValIle 399  
QY 1186 CACATCGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1245  
DB 400 HisThrIleAlaGlyIleValIleValIleValIleValIleValIleValIleValIle 419  
QY 1246 GCGCAACGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305  
DB 420 GlyIleGlyIleValIleValIleValIleValIleValIleValIleValIleValIle 439  
QY 1306 AGCCAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1365  
DB 440 SerTyrLysGlyArgProGlyIleValIleValIleValIleValIleValIleValIle 459  
QY 1366 GCGGAGAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1419  
DB 460 GluIleSerPheArgSerGlyIleValIleValIleValIleValIleValIleValIle 479  
QY 1420 GACGCGCGAG-----ACCTGACAGAGCTGTAAGAGCTGTTCGCGAGAGAGAGAG 1473  
DB 480 AspLysIleLeuTyrProLeuThrSerLeuArgSerIleValIleValIleValIleVal 499  
QY 1474 CAG 1476  
DB 500 Gln 500

RESULT 4  
US-08-375-510-1  
Sequence 1, Application US/08375510  
Patent No. 5576421  
GENERAL INFORMATION:  
APPLICANT: Satco, Atsushi  
APPLICANT: Nakagawa, Hideo  
APPLICANT: Nakata, Atsuo



```

1  TITLE OF INVENTION: HIV ANTIGEN
2  NUMBER OF SEQUENCES: 4
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE: Birch, Stewart, Kolach and Birch
5  STREET: P.O. Box 747
6  CITY: Falls Church
7  STATE: Virginia
8  COUNTRY: U.S.A.
9  ZIP: 22040-0747
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.2
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/375,510
17 FILING DATE:
18 CLASSIFICATION: 435
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US/07/985,949
21 FILING DATE: 04-DEC-1992
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Svensson, Leonard R.
24 REGISTRATION NUMBER: 30,330
25 REFERENCE/DOCKET NUMBER: 216-309P
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (703) 241-1300
28 TELEFAX: (703) 241-2848
29 TELEX: 248345
30 INFORMATION FOR SEQ ID NO: 1:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 500 amino acids
33 TYPE: amino acid
34 STRANDEDNESS: single
35 TOPOLOGY: linear
36 MOLECULE TYPE: protein
37 HYPOTHEetical: NO
38 ANTI-SENSE: NO
39 ORIGINAL SOURCE:
40 ORGANISM: Human immunodeficiency virus type 1
41 US-08-375-510-1

```

```

DB      460 GIJGluserPharqPheIylGluInuThrThrProseGlnlyGlnInuProIle 4
QY      1420 GACCGGAG-----ACCTGACCAAGCTGTAAGAGCTGTGGCAACGACCCCTGAGC 1
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      480 ABpLyGlnLeuTyPProbleuIaSerIeuArgSerIeuPheGlySerAspProSerSer 4
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      1474 CAG 1476
      |||
DB      500 Gln 500

RESULT 5
US-08-487-657-1
; Sequence 1, Application US/08487657
; Patent No. 5834267
; GENERAL INFORMATION:
; APPLICANT: Saito, Atsuehi
; APPLICANT: Sinagawa, Hideo
; APPLICANT: Nakata, Atsuo
; TITLE OF INVENTION: HIV ANTIGEN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch and Birch
; STREET: P. O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,657
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,510
; FILING DATE: 18-JAN-1995
; APPLICATION NUMBER: US/07/985,949
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 216-309P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; US-08-487-657-1

Alignment Scores:
Pred. No.: 6.97e-130 Length: 500
Score: 2249.00 Matches: 421
Percent Similarity: 92.22% Conservative: 41
Best Local Similarity: 84.03% Mismatches: 29
Query Match: 81.13% Indels: 10
DB: 2 Gaps: 5

```

[illegible]

Db 360 AlaArgValIleuLaIaGluIaIaMetSerGlnValThraAsnProAlaThrIleMetIleGln 379  
QY 1126 AAGAGCACTTCAAGGAGCCCGGCGCATGCTCAAGTCTTCAACTGCGGCAAGAGGCG 1185  
Db 380 LysGlyAAsnRheArgAsnGlnArgLysThrValLysSerPheAsnCySgIyLysGlnGly 399  
QY 1186 CACATGCGCCGCAACTGCGGCGCGCGCGCAAGGCGCTGGAAGTGCAGGCGGAGAG 1245  
Db 400 HsIleAlaLysAAsnCySgIyAAsnRheArgLysValLysGlyLysGlyLysGln 419  
QY 1246 GCGCAACAGTGAAGAGTGCAGCGGCGGCGGCAAGTCTTGGGCAAGATCTGAGCC 1305  
Db 420 GlyHisGlnMetLysAAsnRheArgLysThrValLysGlnAlaAsnRheLeuGlyLysIleTrpPro 439  
QY 1306 AGCCCAAGAGGCGCGCGCGGCAACTTCTGCGAGAGCGCGCGGCGGCGGCGCGCGCG 1365  
Db 440 SerHisLysGlyLysArgProGlyAAsnRheLeuGlnSerArgProGlyProThrAlaProPro 459  
QY 1366 GCGGAGAGCTTCCGCTTC-----GAGGAGACACCGCGCGGCAAGAGGAGAGAGAG 1419  
Db 460 GlnGlnSerPheArgPheGlnGlnGlnThrThrProSerGlnLysGlnGlnProIle 479  
QY 1420 GACCGCGAG-----ACCTGACCAAGCTGAGAGAGCTGTCGCAACGACCGCTGAGC 1473  
Db 480 AspLysGlnLeuTyrProLeuAlaSerLeuArgSerLeuPheGlySerAspProSerSer 499  
QY 1474 CAG 1476  
Db 500 Gln 500  
RESULT 5  
US-09-309-572-16  
/ Sequence 16, Application US/09309572  
/ Patent No. 6440730  
/ GENERAL INFORMATION:  
/ APPLICANT: Heinrich-Pette-Institut  
/ TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV  
/ FILE REFERENCE: P50489  
/ CURRENT APPLICATION NUMBER: US/09/309,572  
/ EARLIER FILING DATE: 1999-05-11  
/ EARLIER APPLICATION NUMBER: DE 198 56 463  
/ EARLIER FILING DATE: 1998-11-26  
/ NUMBER OF SEQ ID NOS: 24  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 16  
/ LENGTH: 500  
/ TYPE: PRF  
/ ORGANISM: Human immunodeficiency virus type 1  
/ OTHER INFORMATION: gag pol:protein  
US-09-309-572-16  
Alignment Scores:  
Pred. No.: 6,92e-130 Length: 500  
Score: 2249.00 Matches: 421  
Percent Similarity: 92.22% Conservative: 41  
Best Local Similarity: 84.03% Mismatches: 29  
Query Match: 81.13% Indels: 10  
DB: 4 Gaps: 5  
US-09-475-704a-3 (1-1479) x US-09-309-572-16 (1-500)  
QY 1 ATGGGCGCGCGCGCGCGCATCTGCGCGGCGGCAAGTGCAGTCTGGAAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerValIleuSerGlyGlyGlnLeuAspLysTrpGlnLysIleArg 20  
QY 61 GCGCGCGCGCGCGGCAAGAGTGTACTAGATGAAGCAAGTCTGCGGCGGCGGCGGCGGAG 120  
Db 21 LeuArgProGlyGlyLysLysGlnTyrLysLeuLysHisIleValIleTrpAlaSerArgLys 40  
QY 121 CTGGAAGATTCGCGCTGGAACCGCGCGCTGCTGGAAGACGAGGCGGCTGCAAGCATC 180  
Db 41 LeuGlnArgPheAlaValaAsnProGlyLeuLeuGlnThrSerGlnLysCySgIyGlnIle 60

QY 181 ATCCGCACTGCACCCCGCTGCAAGACCGGCGAGGAGAGTGAAGACCTTTCAC 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnLeuArgSerLeuTyrAAs 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAGAGAGATGAGGTCCGCGACACCAAGAGGCG 300  
Db 81 ThrIleAlaValLeuTyrCySgIyAAsnRheArgLysValLysGlnAlaAspThrLysGlnAla 100  
QY 301 CTGGAACAGATCGAGAGGAGGAGCAAGATGCGGCAAGATTCAGAGCGCGAGGCG 360  
Db 101 LeuAspLysIleGlnGlnLysGlnLysSerLysLysValAlaGlnIleAla---Ala 119  
QY 361 GCGGCAAGGCG-----AAGTGAACCAACTTACCCATGCTGCGAAGTGCAG 411  
Db 120 AlaAspThrGlyAAsnAAsnSerGlnValSerGlnAAsnTyrProIleValGlnAAsnLeuGln 139  
QY 412 GCGCAAGTGTGCACAGGCGCATCAGCGCGCGGCGGCGGCGGCGGCGGCGGATC 471  
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAAsnAlaTrpValLysValAl 159  
QY 472 GAGGAGAGGCGCTTACGCGCGAGGTGATCCCATGTTTACCGCGCTGAGCGAGGCGCG 531  
Db 160 GlnGlnLysAlaPheSerProGlnValIlePrometPheSerAlaLeuSerGlnGlyAla 179  
QY 532 ACCCGCGAGGACCTGAACACAGATGTTGAACACCGTGGCGGCGGCGGCGGCGGCG 591  
Db 180 ThrProGlnAAspLeuAAsnThrMetLeuAAsnThrValGlyLysIleGlnAlaIaMetGln 199  
QY 592 ATGCTGAAGGACACCATCAAGAGAGGCGCGGCGGAGTGGAGCGGCGTGAACCGCGGAC 651  
Db 200 MetLeuLysGlnThrIleAAsnGlnGlnAlaAlaGlnTrpAspArgLeuHisProValHis 219  
QY 652 GCGGCGCGCGCGCGCGCGCGGCGGAGTGCAGAGCGCGCGGCGGCGGCGGCGGCGGCG 711  
Db 220 AlaGlyProIleAlaProGlyGlnMetArgGlnProArgGlySerAspIleAlaGlyThr 239  
QY 712 ACCAGACCTTGAAGAGAGATGCGCTGATGACAGCAACCGCGCGGCTCCCGTGGCG 771  
Db 240 ThrSerThrLeuGlnGlnIleGlyTyrMetThrHisAAsnProProIleProValGly 259  
QY 772 GACATCTCAAGCGGCGGATCATCTCGGCGCTGGAACAGATCGTGGAGTGTACGCGCC 831  
Db 260 GlnIleTyrLysArgTrpIleIleLeuGlyLeuAAsnLysIleValArgMetLysSerPro 279  
QY 832 GTGACATCTTGAACATCAAGAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 891  
Db 280 ThrSerIleLeuAAspIleArgGlnGlyProLysGlnProPheArgAspTyrValaAspArg 299  
QY 892 TTCTTCAGAACCTTGCAGCGCGGCGGAGACACCGGAGGTGAAGAATGTGATGACCGAC 951  
Db 300 PheTyrLysThrLeuArgAlaGlnGlnIleAlaSerGlnGlnValLysAAsnTrpMetThrGln 319  
QY 952 ACCGTGCTGTGCAGAAAGCGCAACCGCGGCTGCAAGACCATCTGCGGCGCTCGGCGCG 1011  
Db 320 ThrLeuLeuValGlnAAsnAAsnProAspCySgIyThrIleLeuLysAlaLeuGlyPro 339  
QY 1012 GCGCGCAGCTGGAAGAGATGTACCGGCTGCGAGGCGGCGGCGGCGGCGGCGGCGG 1071  
Db 340 GlyAlaThrIleGlnGlnIleMetCethrAlaCySgIyGlnValGlyProGlyHisLys 359  
QY 1072 GCGCGCGTGTGCGCGAGGCGGATGAGCGGCGGCAAC-----ACCAAGCTGATGTCAG 1125  
Db 360 AlaArgValaLysAlaGlnAlaMetSerGlnValaIleAsnProAlaThrIleMetIleGln 379  
QY 1126 AAGAGCACTTCAAGGAGCCCGGCGCATGCTCAAGTCTTCAACTGCGGCAAGAGGCG 1185  
Db 380 LysGlyAAsnRheArgAsnGlnArgLysThrValLysSerPheAsnCySgIyLysGlnGly 399  
QY 1186 CACATGCGCGCGCAACTGCGCGCGCGCGCAAGAGGCGTGTGAAGTGCAGGAGAG 1245  
Db 400 HsIleAlaLysAAsnCySgIyAAsnRheArgLysValLysGlyLysGlyLysGln 419

QY	1246	GGCCACCGATGAGAGATGTGACCGAGCGCCAGGCGCAACTCTCGGAGAAATCTGGCCC	1305
Db	420	GLYHSGIMETLYAARPCTYETHCDLARGELINALAANPHELEUGLYLLETTRP	439
QY	1306	AGCCCAAGGGCGCCGCCCGCACTCTCTGAGAGCGCGCCCGAGCCAGCCGCCGCCGCC	1365
Db	440	SEKTIATYSGIYATGPRGILYAASPHLEUENINSEATGPRGGLINPOTHTALAPROPRO	459
QY	1366	GCCGAGAGCTTCGCGCTTC-----GAGGAGACCAACCCCGGCGCAGAGCAGAGAGCAAG	1419
Db	460	GIUQINSEIRPHEALGHPNEGILGELIUNHTRHTRHPROSERGLNLYSGINLIP	479
QY	1420	GACCCGCGAG-----ACCTGACCAAGCCTGGAAGCGCCTGTCGGCAACGACCCCGCTGAGC	1473
Db	480	ASPIYSGIUNLEUTYRPROLEUALSERLEUNRGSRLEUNPHEGLYSERASPPROSESR	499
QY	1474	CAG	1476
Db	500	GLN	500

RESULT 7  
US-09-718-096-16  
; Sequence 16, Application US/09718096  
; Patent No. 6500767

```

: GENERAL INFORMATION:
: APPLICANT: Von Laer, Melke-Dorothee
: TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
: FILE REFERENCE: 35-195

```

;; CURRENT APPLICATION NUMBER: US/09/718, 096  
;; CURRENT FILING DATE: 2000-11-22  
;; PRIOR APPLICATION NUMBER: DE 19856463.5  
;; PRIOR FILING DATE: 1998-11-26

PRIOR APPLICATION NUMBER: EP 99250415.9  
 PRIOR FILING DATE: 1999-11-25  
 PRIOR APPLICATION NUMBER: US 09/309,572  
 PRIOR FILING DATE: 1999-05-11

```

:
: NUMBER OF SEQ ID NOS: 26
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 16
:
: LENGTH: 500
:

```

```

1 NAME: FRI
2
3 ORGANISM: Human immunodeficiency virus type 1
4
5 FEATURE:
6
7 OTHER INFORMATION: gag polypolprotein
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035

```

Alignment Scores: 6.92e-130 Length: 5000  
Pred. No.: 03-03-718-030-16

Percent Similarity:	92.22%
Best Local Similarity:	84.03%
Query Match:	81.13%
Matches:	42
Conservative:	29
Mismatches:	21
Indels:	10

US-09-475-704A-3 (1-1479) x US-09-718-096-16 (1-500)

[illegible]

Db	81	ThrllealavallentryCyvalhlsgrnagllleapvallylsapthrrylsgluala	100
Qy	301	CTGCAACAAGATTCGAGAGAGGACGAAACAAGTCCGACCAAGATCCAGAGGCCGAGGCC	360
Db	101	Leuaperysllleaglulglulgnlansylsserlysllyslalaglnlnla---Ala	119
Qy	361	GCCGCAAGGCC-----AAGGTAGCCGGAACCTAACCCATGTGTGCAGAACTGTGCAG	411
Db	120	Alaaphrtrgylasnasnsertlnvalserglnmentyrprollevalglnabnleugln	139
Qy	412	GGCCGATGTGTGCACACGAGCATATGACGCCCGGACCCGTAAGCGCTGGGTGAAGGTATC	471
Db	140	Glylmetvalhlsgrlnalaleserprokgrthrrleuabnla1trvalylsvala1	159
Qy	472	GAGGAAAGGCGCTTCAAGCCCGAGGTGATCCCATGTTCAACGCCCTGAGCGAGCGCC	531
Db	160	GluGlulysalapheserProglulvalleprometPheSerAlaleusertglulyla	179
Qy	532	ACCCCGGACGCGAACAAGATGTGAACACCGTGGCGGCGACACGAGCGCCATGTGAC	591
Db	180	ThrrproglabpneantthrrMetleuansnhrvalglyglhlsgrlnalalewetgln	199
Qy	592	ATGCTGAAGACACCATCAACGAGAGGCGCGCGAGTGGACCGCGTCAACCCCGTGCAC	651
Db	200	MetleuysrglulThrllleasnglulgnlualaleglultrpaarxglueh1aprovalhls	219
Qy	652	GCGGCGCCCATATCGCCCGCGCAATGGGAGCGCCGCGGACGAGATAGCGCGGAC	711
Db	220	Alaqlprollealaproglglylmetcargluproahglysearbllela1elythr	239
Qy	712	ACGACACCTTCGAGAGACAGATCGCTGTAGTACCAAGCAACCCCGCATCCCGTGGGC	771
Db	240	ThserthrleuglnglulgnlleglyTrpmetThrhl1aenproprolleProValgly	259
Qy	772	GACATCTAACAGCGGTGATCATCTGTGGCTGTAAACAAGTCGTGGATGTACAGCCC	831
Db	260	Glu1elyrlylsaygrlplelleleuglyleuansylleval1argmetyserpro	279
Qy	832	GTGACATCTCGGACATCAACGAGAGGCGCCCAAGAGCGCTCCGCACTAGTGAACGC	891
Db	280	Thserl1leuaprllearglngllyProlysulupropheahgabyrlyValAsparg	299
Qy	892	TTCTTCAAGACCTTCGCGCGCGAGACAGACCCAGAGGTGAAGAACTGATGACCGAC	951
Db	300	Phe1ryrlysrThrleuaglalaglnglnalaserlnglulvallylsanttrpmetthrclu	319
Qy	952	ACCGTCGTGGGACGAGACGCGCAACCCCGACGTGCAAGCAATCCMGAGCGCTGTGGGCC	1011
Db	320	Thrrleuval1glnabn1a1aenproaspysylsnr1leuulysalaleugllypro	339
Qy	1012	GCGCGACCTTCGAGAGATGATGACCGCTTCGAGAGCGGTGGCGCGCCAGCCACAG	1071
Db	340	GlyAla1thrleuglnglulmetMetThAla1cysglnllyal1glylproglhlslys	359
Qy	1072	GCCCGCGTGTGGCGCGAGGAGTGAAGCCAGGCCAAC-----ACCAGCGTATGATGAG	1125
Db	360	Alaargyal1leualaglu1alameSerGlnvalThrsanproalatrlllele1leGln	379
Qy	1126	AAGAGCAACTTCAAGAGGCGCGCGCGCGATGTCAAGTGTCAACTCGGCAAGAGAGGC	1185
Db	380	LyselglyshenhearrgaenlnarlglylsThrVallyscyrheanancysgllysglulgly	399
Qy	1186	CACATCGCCCGCAACTCGCGCGCCCGCGCAAGAGGCGTGTGAAGTGGCGGACGAG	1245
Db	400	H1rlleal1a1ylasbancysarglalaproarlglylsel1Cys1trp1lyscysgllylsglu	419
Qy	1246	GCGCAACAAGTGAAGGTGACACGAGCGCGACGCGCAACTTCCTGGGCAAGATGTGGCCC	1305
Db	420	GlyhlsgrlmetylsabyCys1rrglunagln1a1aenpheuaglyls11etrrpro	439
Qy	1306	AGCCACAAGGCGCGCCCGCAACTTCTTCAGAGCGCGCCGAGCCACCGCCCCCCC	1365

Db 440 SerHieYsGlyArgProGlyAenPheLeuInSerArgProGluProThrAlaProPro 459  
QY 1366 GCCGAGAGCTTCGGCTTC-----GAGAGACACCCCCCGCCGAGGAGGAGAGCAG 1419  
Db 460 GluGlnSerHeArgPheGlyGluGlnThrThrProSerGlnYsGlnGluProIle 479  
QY 1420 GACCGCGAG-----ACCTGACACGAGCTGAGAGCTGTCGCAAGAGAGAGAGGAGG 1473  
Db 480 ArgLysGlnLeuArgProLeuAlaSerLeuArgSerLeuPheGlySerPheProSerSer 499  
QY 1474 CAG 1476  
Db 500 Gln 500  
RESULT 8  
US-09-952-060-35  
Sequence 35, Application US/09952060  
GENERAL INFORMATION:  
APPLICANT: Emili, Emilio A.  
APPLICANT: Youli, Rama  
APPLICANT: Bec, Andrew J.  
APPLICANT: Chen, Ling  
APPLICANT: Kaslow, David C.  
APPLICANT: Shiver, John W.  
APPLICANT: Toner, Timothy J.  
APPLICANT: Casimiro, Danilo R.  
TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV-1 GAG, POL, NEF AND  
FILE REFERENCE: 20747Y  
CURRENT APPLICATION NUMBER: US/09/952, 060  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: PCT/US01/28861  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/317, 814  
PRIOR FILING DATE: 2001-09-07  
PRIOR APPLICATION NUMBER: 60/279, 056  
PRIOR FILING DATE: 2001-03-27  
PRIOR APPLICATION NUMBER: 60/233, 180  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 35  
LENGTH: 1350  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Codon optimized gag-1A pol fusion  
US-09-952-060-35  
Alignment Scores:  
Pred. No.: 1,44e-129 Length: 1350  
Score: 2245.00 Matches: 419  
Percent Similarity: 92.00% Conservative: 41  
Best Local Similarity: 83.80% Mismatches: 32  
Query Match: 80.99% Indels: 8  
DB: 4 Gaps: 4  
US-09-475-704a-3 (1-1479) x US-09-952-060-35 (1-1350)  
QY 1 ATGGGGGGCGGGCGGCGCATCTGCGGCGGCGGCAACGCGCCGAGGAGCGCATCGCC 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyLeuAspLysTrpGlnLysIleArg 20  
QY 61 CTGCGCGCGCGGCGGCAAGAGTGCTCATGATGAGACACCTGCTGTGGCCAGCCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysLysTrpLysLeuLysHieValTrpAlaSerArgGlu 40  
QY 121 CTGGAAGAGTTGCGCTGAAACCCCGGCTGTGCGAGACGAGGAGGCTGCAAGATC 180  
Db 41 LeuGlnArgPheAlaValAsnProGlyLeuLeuGlnThrSerGlnGlyCysArgGlnIle 60

QY 181 ATCCGCGAGCTGCAACCCCGCTGACAGCCGCGAGGAGGAGGCTGTTCAAC 240  
Db 61 LeuGlyLeuLeuInProSerLeuGlnThrGlySerGlnLysLeuArgSerLeuTrpAsn 80  
QY 241 ACCGTGCGCACCTGTACTGCGTGACGAGAGATGAGGTCCGCAACCAAGAGAGCC 300  
Db 81 ThrValAlaThrLeuArgCysValHieGlnLysIleAspValLysAspThrLysGluAla 100  
QY 301 CTGGAAGAATCGAGAGAGAGAGCAACAAAGTCCGACAGAAATCCAGAGCGCCAGGCC 360  
Db 101 LeuGlnLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120  
QY 361 -----GCCGACAGAGGCGAAGTGAGCGAAGTACCCATCGTGCGAAGCGCGG 414  
Db 121 GlyThrGlyAsnSerSerGlnValSerGlnAsnTrpProIleValGlnAsnLeuGlnGly 140  
QY 415 CAGATGTCACACGAGCGCATACAGCCCGCACCTGTAAGCGCTGAGTGAAGTATCGAG 474  
Db 141 GlnMetValHieGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValGlu 160  
QY 475 GAGAGGCTTCAGCCCGGAGGTGATCCCATGTTACCGCCCTGAGCGGCGCACCC 534  
Db 161 GlnLysAlaPheSerProGlnValIleProMetPheSerAlaLeuSerGlnGlyAlaTrp 180  
QY 535 CCCGAGACCTGAAACAGATGTTGAACACCGTGGGCGGCACCGAGCGCGCATGACATG 594  
Db 181 ProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHieGlnAlaAlaMetGlnMet 200  
QY 595 CTGAAGAGACCATCAACGAGAGGCGCGGAGTGAGACCGGCTGACCGCCGTCAGCC 654  
Db 201 LeuLysGlnThrIleAsnGlnGlnAlaIleGlnTrpAspArgLeuHieProValHieAla 220  
QY 655 GGGCCCATGCGCCCGGCGGAGATGCGCGAGCCCGCGGCGGAGCATCGCCGCGCACCC 714  
Db 221 GlyProIleAlaProGlyGlnMetArgLysProArgGlySerAspIleAlaGlyTrpThr 240  
QY 715 AGACCCCTGACAGAGAGATGCGCTGGATGACAGACACCCCGCATCCCGGCGGCGAC 774  
Db 241 SerThrLeuGlnGlnGlnIleGlyTrpMetCnTrpAsnProProIleProValGlyGlu 260  
QY 775 ATCTACAAAGCGGTGATCATCTGCGGCTGGAACAAGATGTCGGATGTACAGCCCGCTG 834  
Db 261 IleTrpLysArgTrpIleIleLeuGlyLeuAsnLysIleValAlaArgMetCysSerProThr 280  
QY 835 AGCATCTGGAACATCAACGAGGCGCCCAAGAGAGCTTCGCGACTGACTGACCGCTTC 894  
Db 281 SerIleLeuAspIleArgGlnGlyProLysGlnProPheArgAspTrpValAspArgPhe 300  
QY 895 TTCAGAACCTGCGCGCGGCGGAGAGACACCGGAGGTGAAGAATCGATGACCGACACC 954  
Db 301 TrpLysThrLeuArgAlaGlnGlnAlaSerGlnGlnValLysAsnTrpMetCnTrpGlnThr 320  
QY 955 CTGCTGTGTCAGAAAGCGCAACCCCGGACTGCAAGACCATCTCGCGCTCGGCGCCGCG 1014  
Db 321 LeuLeuValGlnAsnAlaAsnProArgCysLysThrIleLeuLysAlaLeuGlyProAla 340  
QY 1015 GCGAGCCTGAGAGAGATGATGACCGCTGCGGAGGCGGTGGCGCCCAACCAAGGCC 1074  
Db 341 AlaTrpLeuGlnGlnMetMetCnTrpAlaCysGlnGlnValGlyProGlyHieLysValAla 360  
QY 1075 GCGGTGCTGGCGGAGGCGATGAGCGAGCGCAACACACAGC-----GTGATGATGCGAAG 1128  
Db 361 ArgValAlaLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMetCnGlnArg 380  
QY 1139 AGCAACTTCAAGGCGCGCGCGGCGCATGCTCAAGTCTTCACTGCGCGCAAGAGGCGCAC 1188  
Db 381 GlnAsnPheArgAsnGlnLysTrpValLysCysPheAsnCysGlyLysValGlnLys 400  
QY 1188 ATGCGCGCGCAATGTCGCGCGCGCGCGCAAGAGGCTGTGGAAGTGCAGGCGGCGGCG 1248  
Db 401 IleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGly 420  
QY 1249 CACCAAGATGAAGAGTGCAGCGAGGCGGCGCAACTTCCTGGGCAAGATCTGGCGCGAC 1308

```

Db      421 HtsgImetLyAspCySaSmGluArGlnAlaAnPheLeuGlyLylIleTrProSer 1440
Qy      1309 CCAAGAGGGCGGGCCCCGGGCACTTCTGCAAGAGCCGCCCGAGCCACCGCCCCCGCGCC 1368
Db      441 HtelyGlyArxPrGlyAsnPhelLeuGlnSerArxPrGluInProThrAlaProProGlu 460
Qy      1369 GAGAGCTTCCGCTTC-----GAGAGAGACACCCCGGCGCAAGACAGAGAGCAAGAGAC 1422
Db      461 GlnSerPheArxPheGlyGluGluInuYsrThrTrProSerGlnuYsgInGluProIleAsp 480
Qy      1423 CCGGAG-----ACCTGACCAAGCGCTGAAAGAGCTGTTCGGCAAGACACCCCTTAAGCAG 1476
Db      481 YsgGluLeuYrProLeuAlaSerLeuArxSerLeuPheGlyAsnAspProSerSerGln 500

RESULT 9
US-08-463-210-8
; Sequence 8, Application US/08463210
; Patent No. 6001977
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAL, Floesia
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,210
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193082
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..512
; OTHER INFORMATION: /note= "gag protein of HTLV-III"

US-08-463-210-8

Alignment Scores:      4.99e-129      Length:      512
Pred. No.:            2235.00      Matches:     424
Score:

```

Percent Similarity:	90.64%	Conservative:	41
Best Local Similarity:	82.65%	Mismatches:	25
Query Match:	80.63%	Indels:	22
DB:	3	Gaps:	7
US-09-475-704A-3 (1-1479) x US-08-463-210-8 (1-512)			
QY	1 ATGGAGCGCGCGCGCAGGATCTGTGGCGCGCGCAGCTGGAGACGCTGGAGAGCGCATCCGC	60	
Db	1 MetcGlyAlaHgaHsaSerValIleuSerGlyGlyGluIleuAspArgTrpGluIlySileanG	20	
QY	61 CTGGCGCGCGCGCGCAGAGAGTGCTATCATGATGAAGCACTGGTGTGGCGCAGCGCGAG	120	
Db	21 LeuArgProGlyGlyLysLysLysLysTrpLysLeuLysShiLleValITrpAlaSerArgGlu	40	
QY	121 CTGGAGAGTTCGCGCTGACACCGCGCGCTGCTGGAGACCAAGGAGGCGCTGCAGACGATC	180	
Db	41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluTrpSerGluGlyCysArgGlnIle	60	
QY	181 ATCCGCGCAGCTGCACCCCGCGCTGCAGACCGCGCAGCGAGAGCTGAAGAGCGTGTCAAC	240	
Db	61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluIleuArgSerLeuLysTrp	80	
QY	241 ACCGTGGCCACCTGTATCTGCGTGCACGAGAAATCGAGTTCGCGCAGACCAAGAGGCC	300	
Db	81 ThrValAlaTrpIleuLysCysValIleGlnArgIleGluIleLysAspThrLysGlnAla	100	
QY	301 CTGGCAAGATTCGAGAGAGAGCAGACAAGAGCGCAGAGAGATTCACAGCGCGAGGCC	360	
Db	101 LeuAspLysIleGluGluGlnIleuAsnLysSerLysLysValAlaGlnIleAla--Ala	119	
QY	361 GCCGACAAGGCC-----AAGTGGACCCAGAACTACCCCATCTGTGCAGAACCTGCAG	411	
Db	120 AlaAspThrGlyHisSerSerGlnValSerGlnAsnLysTrpProIleValGlnAsnIleGln	139	
QY	412 GCGCAGATGTGTGCACACGAGGCATACGCGCGCGCAGACCTGCAGACGCTGGGTGAAGTATC	471	
Db	140 GlyIleMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValAla	159	
QY	472 GAGAGAAAGGCGTTCAGCGCGCGAGGTATCCCATGTTCACCGCGCTAGCGAGAGCGCC	531	
Db	160 GluIleLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluLysAla	179	
QY	532 ACCCCCAAGACCTGAAACAGATGTTGAACCGGTGGCGGCCACAGCGCGCATGCAG	591	
Db	180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGln	199	
QY	592 ATGTGTAAGGACCATCAACGAGAGAGCGCGAGTGGAGACCGCGTGCACCGCGTGCAC	651	
Db	200 MetLeuLysGluTrpIleValGlnGlnAlaIleGluTrpAspArgValHisProValHis	219	
QY	652 GCCGCGCCCATCGCGCGCGCAGATCGCAGCGCGCGCGCAGCGCAGCATCGCGCGCACCC	711	
Db	220 AlaGlyProIleAlaProGlyGlnMetArgGluProArgLysSerAlaIleAlaGlyThr	239	
QY	712 ACCAGGACCTTCGAGAGAGAGATGCTGTGATACACGAAACCCCGCATCTCCCGTGGCC	771	
Db	240 ThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnProProIleProValGly	259	
QY	772 GACATCTACAGCGGTGATCATCTGTGGCGCTGAACAAGATCGCGGATGTACAGCGCCC	831	
Db	260 GluIleTrpLysArgTrpIleIleLeuGlyLeuAsnLysIleValAlaGlyMetLysSerPro	279	
QY	832 GTAGAGATCTTGACATCAACAGAGGCGCCCAAGAGGCGCTTCGCGCATACGTGAGCGC	891	
Db	280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTrpValAspArg	299	
QY	892 TTCTTCAAGCCCTGCGCGCGCGCGAGAGACCAAGAGGTGAAGACTGGATACCGAC	951	
Db	300 PheLysTrpThrLeuArgAlaGluGlnAlaSerGlnIleValLysAsnTrpMetThrGlu	319	
QY	952 ACCCTGTGTGACGAAGCGCAACCCCGCATGTGAAGACATCTCGCGCGCTTCGGAGCCC	1011	









```

Db      220  AAGAGTProleuLeuProGluMetArgLysProArgLysSerAspLeuAlaGlyThr 239
QY      712  ACCAGACCCCTGCAGAGACGATGCGCTGGATACACAGACCCCGCCATCCCGTGGGC 771
Db      240  ThrSerThrLeuGlnGlnGlnGlyTyrMetThrAsnAsnProProIleProValGly 259
QY      772  GACATGTACAAGGGGTGGATCATCTGGGGCTGAAGAAGTCTGGGGATGTACAGCCC 831
Db      260  GGuileTyrLeuArgTyrPheIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279
QY      832  GTGAGCATCTGCAGATCAAGCGGGCCCGACAGAGCCCTTCCGCACTACGTGACCGC 891
Db      280  ThrSerIleLeuAsnLysLeuArgGlnGlyProGlyGluProPheArgAspTyrValAspArg 299
QY      892  TTCTTCAAGACCTTGGCGCCCGACAGACACCCAGAGGTGAAGATCGATGATGACCGAC 951
Db      300  PheTyrTyrThrLeuArgLysIleGlnGlnAlaSerGlnGlnValLysAsnTyrMetThrGln 319
QY      952  ACCCTGTGTGAGAACCGCAACCCCGACTGCAACCATCTGGCGCGCTCTCGGGCCGC 1011
Db      320  ThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyPro 339
QY      1012  GCGCGACGCTGGAGAGATGATGATGACCGCTGCGCAGGGCGTGGGGCCCGACACAA 1071
Db      340  AlaAlaThrLeuGlnGlnGlnMetThrAlaCysGlnGlyValGlyGlyProGlyHisLys 359
QY      1072  GCGCGCGTGTGGCGCGAGCGATGACCGC--GCCAACAC--ACCGTATGATGACG 1125
Db      360  AlaArgValLeuValGlnAlaMetSerGlnValThrAsnThrIleMetMetGln 379
QY      1126  AAGAGCACTTCAAGGGCGCCCGCATCTGCAAGTGTCTTCACTGGCGCAAGAGGGGC 1185
Db      380  ArgGlyAsnPheArgAsnGlnArgLysMetValLysCysPheAsnGlyLysGlnGly 399
QY      1186  CACATGCGCCGCACTGCGCGCGCCCGCCGCAAGAGGGCTGTGAAGTGGCGCAAGAG 1245
Db      400  HisThrAlaArgAsnGlyArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGln 419
QY      1246  GCGCACACATGAAGATGACGACCGACCGCGCGCAAGCTTCTGGGGCAAGATGCGGCC 1305
Db      420  GlnHisGlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTyrPro 439
QY      1306  AGCCCAAGAGGGCGCCCGCGCAAC----- 1329
Db      440  SerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProPro 459
QY      1330  TTCCTGCAGAGCGCGCCGACCGACCGCGCCCGCGCGAGAGCTTCCGCG-----TTC 1383
Db      460  PheLeuGlnSerArgProGluProThrAlaProProGlnGlnSerPheArgSerGlyVal 479
QY      1384  GAGGAGACCAACCCCGCGCAAGAGACGAGAGACGACCGCGAG-----ACCTGAGC 1437
Db      480  GlnThrThrThrProProGlnLysGlnLysProIleAspLysIleLeuTyrProLeuThr 499
QY      1438  AGCCGGAAGAGCTGTTCGCAACGACCCCTGAGGCGAG 1476
Db      500  SerLeuArgSerLeuPheGlnAsnAspProSerSerGln 512

```

```

/ TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV-1-GAG, POL, NEF AND
/ TITLE OF INVENTION: MODIFICATIONS
/ FILE REFERENCE: 20747Y
/ CURRENT APPLICATION NUMBER: US/09/952,060
/ CURRENT FILING DATE: 2001-09-14
/ PRIOR APPLICATION NUMBER: PCT/US01/28861
/ PRIOR FILING DATE: 2001-09-14
/ PRIOR APPLICATION NUMBER: 60/317,814
/ PRIOR FILING DATE: 2001-09-07
/ PRIOR APPLICATION NUMBER: 60/279,056
/ PRIOR FILING DATE: 2001-03-27
/ PRIOR APPLICATION NUMBER: 60/233,180
/ PRIOR FILING DATE: 2000-09-15
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 33
/ LENGTH: 493
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: tpa-gag fusion open reading frame
US-09-952-060-33
Alignment Scores:
Pred. No.: 3,79e-121 Length: 493
Score: 2106.00 Matches: 394
Percent Similarity: 92.29% Conservative: 37
Best Local Similarity: 84.37% Mismatches: 28
Query Match: 75.97% Indels: 8
DB: 4 Gaps: 4
US-09-475-704A-3 (1-1479) x US-09-952-060-33 (1-493)
QY      100  CTGGTGTGGGCGGACCGCGAGCTGGAAGAGTTGCGCCCTGAGACCCCGGCTGTGAGACC 159
Db      27  ILevalTrrpAlaSerArgGluLeuGlnArgPheAlaValAsnProGlyLeuLeuGlnThr 46
QY      160  AGCGAGGCTGAGAGATGATCATCGCGAGCTGACACCCCGCTGAGACCGCGAGCGAG 219
Db      47  SerGlnGlyCysArgGlnIleLeuGlnGlnLeuGlnProSerLeuGlnThrGlySerGln 66
QY      220  GAGCTGAAGAGCTGTTCACACCGTGGCGCACCTGTACTGCGTGCAGACGAGAGATCGAG 279
Db      67  GlnLeuAsnSerLeuTyrAsnThrValAlaThrLeuTyrCysValHisGlnLysIleAsp 86
QY      280  GTCCCGACACCAAGAGGCGCTTGCACAGATCGAGAGAGACCAAGATGCGACGAG 339
Db      87  ValLysAspThrLysGlnAlaLeuGlnLysIleGlnGlnGlnLysAsnLysSerLysLys 106
QY      340  AAGATCCAGACGCGCGAGGCC-----GCCGCAAGGGCGAGGTGAGCCGACCAATACCCC 393
Db      107  LysAlaGlnGlnAlaAlaIleGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrPro 126
QY      394  ATGCTGCAGAACTCTCAGAGGCGCAGATGTGACACGAGCCATCAGGCGCGCACCTGAA 453
Db      127  IleValGlnAsnLeuGlnGlnGlnMetValHisGlnAlaIleSerProArgThrLeuAsn 146
QY      454  GCGTGGTGAAGGTATCGAGAGAGGCTTACGCGCCCGGAGGTGATCCCATGTTACCC 513
Db      147  AlaTrrpValLysValIleGlnGlnLysLysAlaPheSerProGlnValIleProMetPheSer 166
QY      514  GCGCTGAGCGAGGGGCGGACCCCGCGAGACCTGGAACAGATTTGAAACCTGTGGCGGCG 573
Db      167  AlaLeuSerGlnGlyAlaThrProGlnAsnPheLeuAsnThrMetLeuAsnThrValGlyLys 186
QY      574  CACGAGCGCGCATGCTGAGAGACACCATCAACGAGAGAGGCGCGCGAGTGGAGC 633
Db      187  HisGlnAlaAlaMetGlnMetLeuLysGlnThrIleAsnGlnGlnAlaIleLysTrrpAsp 206
QY      634  CGGCTGACACCCGCTGACAGCGCGCGCCATCGCGCCCGCGCGAGATCGCGAGCGCGCG 693
Db      207  ArgLeuHisProValHisAlaGlyProIleAlaProGlyGlnMetArgGlnProArgGly 226

```

```

/ RESULT 12
/ US-09-952-060-33
/ Sequence 33. Application US/09952060
/ Patent No. 6733993
/ GENERAL INFORMATION:
/ APPLICANT: Emilio A.
/ APPLICANT: Youli, Rima
/ APPLICANT: Betz, Andrew J.
/ APPLICANT: Chen, Ling
/ APPLICANT: Kaslow, David C.
/ APPLICANT: Shiver, John W.
/ APPLICANT: Toner, Timothy J.
/ APPLICANT: Casimiro, Danilo R.
/ TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS

```

QY	694	AGCGAATGCGCGGCACACACACACCGCTTCAGAGAGAGATGGCTGGATGACAGCAAC	753
Db	227	SeapApIIeAlaGlyThrThrSerThrLeuGlnGlnIleLeuTyrPectinAseAse	246
QY	754	CCCCCATCCCGCTGGGCGACATCTCAACACGGTGATCATCTCGGCGCTGAACAAGATC	813
Db	247	ProPoiIleProValGlyIleTyrLysArgTrrIleIleLeuGlyAseAseIle	266
QY	814	GTGGCGAATGTACAGCCCGGTGAGATCTCTGAGATCAAGAGGGCCCCAAGAGCCCTTC	873
Db	267	ValArgMetTyrSerProThrSerIleLeuAspIleArgGlnIleProLysGluProPhe	286
QY	874	CGCGACTACGTGACCGCTTCTTCAAGACCCCTCGCGCGCGAGAGAGACCCAGAGGGG	933
Db	287	ArgAspTyrValAlaAspArgPheTyrLysThrLeuArgAlaGlnIleAlaSerGlnIleVal	306
QY	934	ANGAATGTGATGACCGACACCCCTGCTGTGTGACAGAACGCCAACCCCGCATGTCAAGACATC	993
Db	307	LysAsenTrrMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIle	326
QY	994	CTGCGCGCTCTGGCGCGCGCGCGCGCGCTGTGAGAGATATGATGACCGCTGCCAGGGCGTG	1055
Db	327	LeuLysAlaLeuGlyProAlaAlaThrLeuGlnGluMetMetThrLysCysGlnIleVal	346
QY	1054	GGCGGCGCCGACGCACAGAGCGCGCGCTGTGCGCGAGCGCATGAGCGAGCGCAACACACAGC	1114
Db	347	GlyGlyProGlyLysIleLysAlaArgValAlaLeuAlaGlnIleMetSerGlnValThrAsenSer	366
QY	1114	-----GTGATGATGACAGAGAGCACTTCAGAGGGCGCGCGCGCATGCTCAAGTGCTTC	1168
Db	367	AlaThrIleMetMetGlnArgGlyAsnPheArgAsnGlnIleArgLysThrValLysCysPhe	386
QY	1168	AACTGCGGGAAGAGAGGGCCACATGCGCGCGCACTGCGCGCGCGCGCGCAAGAGGCTTC	1228
Db	387	AsnCysGlyLysValGlyIleLysIleAlaLysAsnCysArgAlaProArgLysLysGlyCys	406
QY	1228	TGGAAGTGCAGAGAGGCGCACCAAGATGAAGAGACTGACACGAGCGCGACGCGCAACTTC	1288
Db	407	TrrLysCysGlyLysGlnGlyLysIleGlnMetCysAspArgAsnGlnIleAlaAsnPhe	426
QY	1288	CTGGGCAAGATTTGGCCCGACGCCACAAAGGGCGCGCGCGCACTTCTGTGAGAGCGCGCC	1344
Db	427	LeuGlyLysIleTrrProSerIleLysGlyArgProGlyAsnPheLeuIleAsnIleArgPro	446
QY	1348	GAGCCCAACGCGCGCGCGCGCGAGAGGCTTCGCGCTTC-----GAGAGAGCAACCCCGCGC	1408
Db	447	GluProThrAlaProProGluLysIleSerPheArgPheGlyGlnIleLysThrThrProSer	466
QY	1402	CAGAAAGCAGAGAGAGAGACCGCGAG-----ACCTTGACACAGCGCTGAAGAGCTGTTC	1458
Db	467	GlnLysGlnGluProIleAspLysGlnLeuTyrProLeuAlaSerLeuArgSerLeuPhe	486
QY	1456	GGCAACGACCCCTGAGCCAG 1476	
Db	487	GlyAsnAspProSerSerGln 493	

RESULT 13  
US-08-127-499A-11  
Sequence 11, Application US/08127499A  
Patent No. 5510264  
GENERAL INFORMATION:  
APPLICANT: VAN ALSTYNE, Diane  
APPLICANT: SHARMA, Lawrence Rajendra  
TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED  
TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

```

?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/127,499A
? FILING DATE: 28-SEP-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: BENT, Stephen A.
? REGISTRATION NUMBER: 29,768
? REFERENCE/DOCKET NUMBER: 51916/102/INBI
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202)672-5300
? TELEFAX: (202)672-5399
?
? TELEX: 904136
?
? INFORMATION FOR SEQ ID NO: 11:
?
? SEQUENCE CHARACTERISTICS:
? LENGTH: 478 amino acids
? TYPE: amino acid
? STRANDEDNESS:
?
? TOPOLOGY: unknown
?
? US-08-127-499A-11

```

Pred. No.:	5,78e-117	Length:	47
Score:	2037.50	Matches:	38
Percent Similarity:	90.81%	Conservative:	47
Best Local Similarity:	81.00%	Mismatches:	33
Query Match:	73.50%	Indels:	11
DB:	1	Gaps:	8

US-09-475-704A-3 (1-1479) x US-08-127-499A-11 (1-478)

[illegible]

QY	599	AGCTCAAGACACACATCAACAGAGAGGCGCCGAGTGGACCGCGTGCAC	651
Db	200	MeLEuLyGdIuThrILeAnGInGInUaIaGInUThraSPaTgVInILaProVaILhIS	219
QY	652	GCCGAGCCCATGCGCCCGCCGAGATGCGGAGCCCGCGGACGACATGCGGACCC	711
Db	220	AlaGlyProILeAlaProGlyGInMeTArgGInProArgILySerAPILeAlaGlyThr	239
QY	712	ACCAGACCCCTGAGAGAGCATGCGCTGGATATACAGCAACCCCGCATCCCGTGGGC	771
Db	240	ThrSerThLeuGInGInGInILeGlyTyrMeTThrAnSProPILeProVaILy	259
QY	772	GACATCTTAAGCGGTGATCATCTTGAGCGCTAAACAAGATCGTGGCGATGTACAGCCCC	831
Db	260	GIuILeTyLybArgTrpILeILeGInGlyLeuAnSyILeVaILaGMeTyTyrSerPro	279
QY	832	GTGAGCATCTTGACATCAAGCAAGGCGCCCAAGAGAGCCCTTCGCGCATAGTGAGCCGC	891
Db	280	ThrSerILeLeuAnPILeArgGInGlyProLyGInProPheArgPeryVaILaPaArg	299
QY	892	TTCTTCACAGACCTTGCGCGCGGAGAGAGACCCAGAGAGGTGAAGATGTAGTACCGAC	951
Db	300	PheTyLybThLeuArgTrpILeILeGInGInILaSerGInGInVaILyAnTPMeTThGIn	319
QY	952	ACCGTGTGTGTCAGAACCCCAACCCCGATGGAAGACATCTCGTCCGCGTTCGAGCCCC	1011
Db	320	ThrLeuLeuVaILeGInMeILaAnPILeProArgCyLybThrILeLeuSyVaILeGlyPro	339
QY	1012	GCGCCGACCTTGAGAGATGATGATGACCGCGCTGCGAGGCGCTGAGCGCGCCCAAGCACAAG	1071
Db	340	AlaAlaThLeuGInGInUeMeTThrILaCyGInGInGlyILyGlyProGInILeLy	359
QY	1072	GCGCGGTGCTGGCGGAGCGATGAGCCGAG---GCCAACAC---AGCGTGAATGATCAG	1125
Db	360	AlaArgVaILeUaILaGInILaMeTSerGInVaILThraSnThraILaThrILeMeTMeGIn	379
QY	1126	AAGACCAATTCAGAGAGGCGCCCGCGCCCATGTCAGAGTCTTCAACTGCGGCAAGAGAGGC	1185
Db	380	ArgGLyAnPheArgAnGInnArgLybMeVaILyCybPheAnCybGlyLybGInGly	399
QY	1186	CACATCGCCCGCACTGCGCGCGCCCGCCGCAAGAGAGGCGTCTGGAATGCGCGGACAGAG	1245
Db	400	HisThraILaArgAnCybArgILaProArgLybGlyCybTrpLybCybGlyLybGIn	419
QY	1246	GCGCCACAGATGAAGAGTGCACCGGCGGAGGCGCAACTTCCTGGCGCAAGATC---TGG	1302
Db	420	GLyILaGInMeLybAspCybTrpGInUaArgGInILaSnPheUeGlyLybILeCybLeu	439
QY	1303	CCGACGACCAAGAGGCGCGCCCGCGCACTTCCTGCAGAGCGCGCCCGAGCCACCGCCCC	1362
Db	440	ProThraArgGInGlyGIn---GlyILePhePhe-----ArgILaAnPILeGInGIn	456
QY	1363	CCCGCGGAGAGCTTCGCGTTCAGAGAGAC---ACCGCGGCGCAAGACGAGAGAC	1416
Db	457	ProHisILaSPhePheArgILaAnPILeGInGInProHisGInGInUaArgILaSer	475
RESULT 14			
US-08-482-847-11			
Sequence 11, Application US/08482847			
Patent No. 5556757			
GENERAL INFORMATION:			
APPLICANT: VAN ALSTYNE, Diane			
TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR			
BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR			
TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF			
NUMBER OF SEQUENCES: 40			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Foley & Lardner			
STREET: 3000 K Street, N.W., Suite 500			
CITY: Washington			
STATE: D.C.			

```

? COUNTRY: USA
? ZIP: 20007-5109
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/482,647
? FILING DATE: 07-JUN-1995
? CLASSIFICATION: 514
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/127,499
? FILING DATE: 28-SEP-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: BENT, Stephen A.
? REGISTRATION NUMBER: 29,768
? REFERENCE/DOCKET NUMBER: 51916/104/INBI
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 672-5300
? TELEFAX: (202) 672-5399
? TELEX: 904136
? INFORMATION FOR SEQ ID NO: 11:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 478 amino acids
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: unknown
?
? US-08-482-847-11
?
Alignment Scores:
Score: No: 5,78e-117 Length: 478
Percent Similarity: 2037.50 Matches: 388
Best Local Similarity: 90.81% Conservative: 47
Query Match: 81.00% MisMatches: 33
DB: Query Match: 73.50% Indels: 11
Gaps: 8
US-09-475-704A-3 (1-1479) x US-08-482-847-11 (1-478)
QY 1 ATGGGCGCCCGCGCGCAGATCTCTGCGCGCGGCAAGTCGACGCCGTGGAGCGATCCGC 60
Db 1 MetGlYAlaArgAlaSerValLeuSerGlyGluLeuAspArgTrpGluTyrIleArg 20
QY 61 CTCGCCCGCCGGCGGCAAGAAGTCATCATATGAAGCACTGTGTGTGGCCACGCCGAG 120
Db 21 LeuArgProGlyGlyLeuLysLeuTybTyrLysLeuYshIleValITrpAlaSerArgGlu 40
QY 121 CTGGAAGAAGTGCGCCCTGAAACC CGCGCCTGTGAGACACGACGAGGGGTGCAAGACATC 180
Db 41 LeuGlnArgPheAlaValAsnProGlyLeuGlnThrSerGluGlyCybArgGlnIle 60
QY 181 ATCCGCGACCTGCACCCCGCCCTGCAGACCGGACGAGAGAGCTGAAGAGCTTTCAAC 240
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGlnLeuArgSerLeuTyrAsn 80
QY 241 ACCGTGGCCACCTGTACTGTCTGTGACGAGAAAGATCCAGGTCGCCCGCACCAAGAGAGCC 300
Db 81 ThrValAlaThrLeuTyrCybValHisGlnArgIleGluTyrIleLysAspThrLysGluAla 100
QY 301 CTGGAACAATGCAGGAGGAGGACGAAGAAGTGCAGACGAGAAGATCCAGACGGCGGAGCC 360
Db 101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 119
QY 361 GCCGACAGAGGC-----AGGTGACCAAACTACCCCATCTGTGCAGAACCTGCAG 411
Db 120 AlaAspThrGlyHisSerSerGlnValSerGlnAsnTyrProIleValIGlnAsnIleGln 139
QY 412 GGCACAATGTCACACAGGCAATCAGCCCGGACCCCTGAAAGCGCTGGAGTGAAGTATC 471
Db 140 GlyGlnMetValHisGlnAlaIleSerSerArgTrpGlnThrLeuAsnAlaTrpValLysVal 159
QY 472 GAAGAAAGGCTTAGCCCCGAGGTGATCCCATTGTTCACCGCCCTGACGAGAGGCGCC 511

```

Db GluGluIysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179  
160  
QY ACCCCCCAGAGCCTGAAACAGATGTTGAACAACCTGTGGCGGCAACGAGCCGCATGCAG 591  
Db ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGln 199  
QY ATGCTGAAGGACACCATCAACGAGAGCGCGGATGGGAGCCGGTGAACCCCGTGCAC 651  
Db MetLeuIysGluThrIleAsnGluGlnAlaAlaGluThrAspArgAlaHisProValHis 219  
200 MetLeuIysGluThrIleAsnGluGlnAlaAlaGluThrAspArgAlaHisProValHis 219  
QY GCCGGCCCATCGCCCCCGGCAAGATCGCGAGCCCCGCGGCAACATCGCCGAGCAC 711  
Db AlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThr 239  
QY ACCAGGACCTCTGCAGAGAGATCGCTGGATATCAACGACCCGCCCATCCCGTGGGC 771  
Db ThrSerThrLeuGlnGlnGlnIleGlyTyrMetThrAsnAsnProProIleProValGly 259  
240 ThrSerThrLeuGlnGlnGlnIleGlyTyrMetThrAsnAsnProProIleProValGly 259  
QY GACATCTCAAGCGGTGGATCATCTCGGGCGCTCAACAAGATCGTGGATGTAACGCCCC 831  
Db GluIleTyrIysAsnGlyPheIleIleLeuGlyLeuAsnIlyIleValArgMetCysSerPro 279  
260  
QY GTAGCATCTCTGCATCAACGAGGCGCCCAAGAGCCCTTCGCGCATACGTGAGCCGC 891  
Db ThrSerIleLeuAspIleArgGlnGlyProIysGluProPheArgAspTyrAlaAspArg 299  
280 ThrSerIleLeuAspIleArgGlnGlyProIysGluProPheArgAspTyrAlaAspArg 299  
QY TTCTTCAAGACCTCGCGCGCGGCAAGAGACCCAGAGAGTGAAGAATCTGATGACGAC 951  
Db PheTyrIysThrLeuAsnArgAlaGlnGlnAlaSerGlnGlnIlyValAsnThrMetThrGln 319  
300 PheTyrIysThrLeuAsnArgAlaGlnGlnAlaSerGlnGlnIlyValAsnThrMetThrGln 319  
QY ACCCTGCTGGTGGAGAACGCGCAACCCCGACTGCAGAACCATCTCGCGCTTCGGGCC 1011  
Db ThrIleuValGlnAsnAlaAsnProAspCysIlyThrIleLeuIlyAlaLeuGlyPro 339  
320 ThrIleuValGlnAsnAlaAsnProAspCysIlyThrIleLeuIlyAlaLeuGlyPro 339  
QY GCGCGCAGCTGAGAGATGATGACCGCTGCGAGGCGTGGGCGGCCCAAGCCACAG 1071  
Db AlaAlaThrLeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisIlyAs 359  
340 AlaAlaThrLeuGlnGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisIlyAs 359  
QY GCCCGCGTGTGGCGGAGGCGAATGAGCCAG---GCCAACAC---AGCGTATGATGACG 1125  
Db AlaArgValIleuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGln 379  
360 AlaArgValIleuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrIleMetMetGln 379  
QY AAGAGCAACTTCAAGGCGCCCGCGCGCATGTCATGATGCTTCACTGCGGCAAGAGGAC 1185  
Db ArgGlyAsnAspThrArgAsnGlnArgIlyMetValIlyCysPheAsnCysGlyIlyLeuGly 399  
380 ArgGlyAsnAspThrArgAsnGlnArgIlyMetValIlyCysPheAsnCysGlyIlyLeuGly 399  
QY CACATGCGCCCGCAATCGCGCGCCCGCGCAAGAGAGGCTGTGGAAGTGGCGGCAAGAG 1245  
Db HisThrAlaArgAsnCysArgAlaProArgIlyIlyGlyCysTyrIlyCysGlyIlyLeuGly 419  
400 HisThrAlaArgAsnCysArgAlaProArgIlyIlyGlyCysTyrIlyCysGlyIlyLeuGly 419  
QY GGCACCAAGATGAAGATGACACGAGGCGCGAGGCGCACTTCTGGGGCAAGATC---TGG 1302  
Db GlyHisIleMetIysAsnArgCysThrGlnIlyGlnAlaAsnPheLeuGlyIlyCysLeu 439  
420 GlyHisIleMetIysAsnArgCysThrGlnIlyGlnAlaAsnPheLeuGlyIlyCysLeu 439  
QY CCCAGCCACAGGGCGCGCCCGGCAACTTCTGCAAGAGCGCGCCGAGCCGAGCCGCC 1362  
Db ProThrArgGluGlyGln---GlyIlePhePhe-----ArgAlaAspGlnSerGlnGln 456  
440 ProThrArgGluGlyGln---GlyIlePhePhe-----ArgAlaAspGlnSerGlnGln 456  
QY CCGCGCGAAGACTTCCGCTTCAAGAGAC---ACCCCGCGAGAGACAGAGAGC 1416  
1363 CCGCGCGAAGACTTCCGCTTCAAGAGAC---ACCCCGCGAGAGACAGAGAGC 1416  
457 ProHisIlePheThrArgAlaAspGlnSerGlnGlnProHisGlnIlyAspArgIlySer 475

```

RESULT 15
US-08-392-794A-2
; Sequence 2, Application US/08392794A
; Patent No. 6025141
;
; GENERAL INFORMATION:
;   APPLICANT: HU, Yu-Wen
;   TITLE OF INVENTION: IMMUNOFLOURESCENCE ASSAY FOR THE DETECTION OF
;   TITLE OF INVENTION: ANTIBODIES USING RECOMBINANT ANTIGENS IN INSOLUBLE FORM
;   NUMBER OF SEQUENCES: 21
;   CORRESPONDENCE ADDRESS:

```

ADDRESSSEE: BROWDY AND NEIMARK  
 STREET: 4419 Seventh Street, N.W.  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/392,794A  
 FILING DATE: 09-DEC-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/164,789  
 FILING DATE: 10-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: NEIMARK, SHERIDAN  
 REGISTRATION NUMBER: 20,520  
 REFERENCE/DOCKET NUMBER: HU-4A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-628-5197  
 TELEFAX: 202-737-3528  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 437 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-392-794A-2

Pred. No.:	4.38e-116	Length:	4
Score:	2023.00	Matches:	3
Percent Similarity:	94.06%	Conservative:	3
Best Local Similarity:	85.94%	Mismatches:	2
Query Match:	72.98%	Indels:	3
DB:	3	Gaps:	6

US-09-475-704A-3 (1-1479) X US-08-392-794A-2 (1-437)

[illegible]



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 14:56:14 ; Search time 202.694 Seconds

(without alignments)  
5644.151 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 2772

Sequence: 1 atggcgccgcgcgcagcat.....acgaccctctgagccagtaa 1479

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

MODEL=frame\_n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO\_SPOOL\_P/US09475704/rnatc.10032005.140220.14808/app\_query.fasta\_1.3342  
-DB=A.GeneSeq.16Dec04 -QFMT=fastaan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biolum62 -TRANS=humana40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=prto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704@cgn.1.1.472@rnatc.10032005.140220.14808 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONCLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.GeneSeq.16Dec04:\*

1: geneeqp1980s:\*\n2: geneeqp1980s:\*\n3: geneeqp2000s:\*\n4: geneeqp2001s:\*\n5: geneeqp2002s:\*\n6: geneeqp2003as:\*\n7: geneeqp2003bs:\*\n8: geneeqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2606	94.0	492	3	AA96943 HIV synch
2	2448	88.3	492	6	AA37600 HIV-1 sub
3	2441	88.1	492	8	ADP02848 wild type
4	2441	88.1	513	8	ADP02847 wild type
5	2436.5	87.9	514	8	ADN36386 Human pro
6	2435	87.8	492	5	AAE18303 Human imm
7	2435	87.8	635	5	AA448948 HIV-1 sub
8	2433	87.4	492	7	ADCT2880 HIV-1 gag
9	2419.5	87.3	1457	8	ADN36406 Human pro
10	2416	87.2	492	7	ADCT2879 HIV-1 gag

11	2416	87.2	494	3	AA869275
12	2415.5	86.9	491	7	ADCT2910
13	2407.5	86.9	491	7	ADCT2874
14	2405.5	86.8	491	3	AA869278
15	2396.5	86.5	491	7	ADCT2913
16	2383	86.0	498	3	AA869276
17	2382	85.9	496	7	ADCT2878
18	2381.5	85.9	487	7	ADCT2875
19	2381.5	85.9	497	7	ADCT2877
20	2344	84.6	508	7	ADCT2875
21	2343	84.5	502	7	AA96944 HIV synch
22	2341.5	84.5	3025	4	AA86196 HIV-1 sub
23	2339	84.0	494	3	AA869271
24	2328.5	84.0	494	7	ADCT2896
25	2301	83.0	500	1	AA80884 HIV-1 gag
26	2297	82.9	500	7	ADCT2788
27	2294.5	82.8	497	2	AA80840 HIV-1 gag
28	2277	82.1	498	8	AD052538
29	2269	81.9	500	6	ABR5448 HIV-1 gag
30	2267.5	81.8	500	3	AA869270
31	2266.5	81.8	505	1	AA80885 HIV-1 gag
32	2266.5	81.8	505	2	AAW72992
33	2266	81.7	500	8	AD052543
34	2263	81.6	500	8	AD052550
35	2262	81.6	498	8	AD052543
36	2260	81.5	508	1	AA869419
37	2259.5	81.5	492	7	AA869268
38	2258	81.5	500	7	ADCT2891
39	2258	81.5	508	5	AA019386
40	2256	81.4	500	2	AAW30562
41	2256	81.4	500	6	AAW30569
42	2256	81.4	500	2	ABR44737
43	2256	81.4	500	6	AA832734
44	2256	81.4	500	7	AD881007
45	2256	81.4	500	7	AD860702

#### ALIGNMENTS

RESULT 1	AA96943	standard; protein; 492 AA.
XX	AA96943;	
AC	AA96943;	
XX		
DT	31-OCT-2000	(first entry)
XX		
DE	HIV synthetic Gag polypeptide.	
XX		
KM	Gag/ expression cassette; antigenic; type C, HIV; Env, synthetic;	
KW	DNA immunization; packaging cell line; antigen presentation.	
OS	Human immunodeficiency virus; type C strain AF110965.	
XX	Synthetic.	
PN	WO200039304-A2.	
XX		
PD	06-JUL-2000.	
XX		
PF	30-DEC-1999;	99WO-US031273.
XX		
PR	31-DEC-1998;	98US-011495P.
XX		
PR	01-SEP-1999;	99US-0152195P.
XX		
PA	(CHIR ) CHIRON CORP.	
XX		
PI	Barnett S, Zur Megede J;	
XX		
DR	WPI; 2000-452401/39.	
XX		
DR	N-PSDB; AA51609.	
XX		
PT	Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env	
	polypeptide and the polypeptide useful for immunizing a mammal especially	





PF 31-OCT-2002; 2002MO-IB04550.  
XX  
PR 31-OCT-2001; 2001ZA-00008978.  
XX  
PA (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.  
XX (UYCA-) UNIV CAPE TOWN.  
XX  
PI Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;  
XX WPI; 2003-430497/40.  
DR N-PSDB; AAD29257.  
XX  
PT New molecules comprising HIV-1 subtype isolate regulatory/accessory  
PT gene, useful for manufacturing a vaccine for treating or preventing HIV  
XX infection.  
XX  
XX Disclosure; Page 91-92; 97pp; English.  
XX  
XX The invention relates to molecules comprising HIV-1 subtype isolate  
XX regulatory/accessory genes (tat, nef and rev genes) and modifications and  
XX derivatives thereof. The invention also provides proteins encoded by such  
XX genes. Sequences of the invention are useful for manufacturing vaccines  
XX for treating or preventing human immunodeficiency virus (HIV) infections.  
XX They are also useful in gene therapy. The present sequence is HIV-1  
XX subtype C isolate Dna22 Gag protein. (Updated on 23-Oct-2003 to  
XX standardise OS field)  
XX  
SQ Sequence 492 AA;  
XX  
Alignment Scores:  
Pred. No.: 1,05e-137 Length: 492  
Score: 248.00 Matches: 461  
Percent Similarity: 97.36% Conservative: 19  
Best Local Similarity: 93.51% Mismatches: 11  
Query Match: 88.31% Indels: 2  
DB: 6 Gaps: 2  
US-09-475-704A-3 (1-1479) x AAE37600 (1-492)  
QY 1 ATGGGCGCCGCGCGAGCATCTCGCGCGCGAGCGCTGGAGCGCATCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGluLeuLeuPyrTrpGluYsIleArg 20  
QY 61 CTGGCGCCCGCGCGCAAGAGTCTACATGATGAACACCTGGTGGCCGACCCGCG 120  
DB 21 LeuArgProGlyGlyLeuYsYbHLeuMetLeuYsHleValTrpAlaSerArgGlu 40  
QY 121 CTGGAGAGATTGGCCCTGAACCCCGGCTGGTGAACAGCGAGGGCTGCAAGCATC 180  
DB 41 LeuGluArgPheHleuLeuAsnProGlyLeuLeuGluThrSerGluGlyYsGlnIle 60  
QY 181 ATCCGCGAGCTGACCCCGCCCTGACAGCCGCGAGAGAGCTGAAGACCTGTTCAAC 240  
DB 61 MetYsGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuYsSerLeuYrAsn 80  
QY 241 ACCGTTGGCCACCTGTAAGTCTGTCGACAGAGATGAGTCCGCGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuYrCysValHleGluYsIleGluValArgAspThrYsGlnAla 100  
QY 301 CTGGCAAGATCGAGAGAGAGAGACAGAGTGCACAGAGATCCAGAGGCGCGAGGCC 360  
DB 101 LeuAspYsIleGluGlnGluGlnAsnYsCysGlnGlnYrThrGlnGlnAlaYsAla 120  
QY 361 GCGCAGCAGAGGAGAGTGAAGCAGAACTACCCCATGTCGACAACTGCAAGGCGCATG 420  
DB 121 AlaAsp---GlyYsValSerGlnAsnYrProIleValGlnAsnLeuGlnGlnMet 139  
QY 421 GTGCAACAGGCGCATGAGCCCGCGCACCTGGAAGCTGGTGAAGTATCGAGAGAG 480  
DB 140 ValHleGlnAlaHleSerProArgThrLeuAsnAlaTrpValYsValIleGluGluYs 159  
QY 481 GCCTTCAAGCCGAGGATGATCCCATGTTCAACGCGCTGAGAGAGGCGCGACCCCGCAG 540

DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGluYrAlaThrProGln 179  
QY 541 GACCTGAACACATGTTGAACACCGTGGCGGCCACAGCGCGCATGCAATGCTGAAG 600  
DB 180 AspLeuAsnThrMetLeuAsnThrValGlyGlnHleGlnAlaAlaMetGlnMetLeuYs 199  
QY 601 GACACCATCAAGAGAGGCGCCCGGAGTGGGACCGGCTGCACCCGTCAGCGCGGCC 660  
DB 200 AspThrIleAsnGluGlnAlaHleGluTrpAspArgValHleProValHleAlaGlyPro 219  
QY 661 ATGCGCCCGCGCAGATGCGCGAGCGCCGCGCAGCAGATGCGCGCAGCAGCAGCACC 720  
DB 220 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGCAAGAGACATGCGCTGATGACAGCAACCCCGCATTCGCGTGGCGCAATGAC 780  
DB 240 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProProIleProValGlyAspIleYr 259  
QY 781 AAGCGGTGATGATCTGGCGCTGGAACAGATCTGCGGATGATGACCGCCGCGAGCATC 840  
DB 260 YsArgTrpIleIleLeuGlyLeuAsnYsIleValArgMetYrSerProValSerIle 279  
QY 841 CTGACATCAAGAGGCGCCCAAGAGACCTTCGCGCATACGATGACCGCTTCTTCAAG 900  
DB 280 LeuAspIleArgGlnGlyProYsGluProPheHleArgAspYrValAspArgPhePheYs 299  
QY 901 ACCCTGCGCGCGCAGCAGACCCAGAGAGTGAAGATCTGATGACCGACACCTGCTTG 960  
DB 300 ThrLeuArgAlaGluGlnAlaThrGlnGluValYsAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGCAAGACCCAGACCCCGCATGCAAGACCATCTCGCGGCTCTGGCGCCGCGCGCAC 1020  
DB 320 ValGlnAsnAlaAsnProAspYsYbHTrpIleLeuArgAlaLeuGlyProGlyAlaThr 339  
QY 1021 CTGAGAGAGATGATGACCGCTGCCAGGCGCTGGCGCGCGCCAGCACAAGCGCCGCTG 1080  
DB 340 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyYrHleYsAlaArgVal 359  
QY 1081 CTGGCCGAGCGATGAGCCAGGCCAACCC---ACCGTATGATGACAGAAAGCAATTC 1137  
DB 360 LeuAlaGluAlaMetSerGlnAlaAsnSerGlyAsnIleMetMetGlnArgSerAsnPhe 379  
QY 1138 AAGCGCCCGCGCGCATGTCAGAGCTTCAACCTCGCGCAAGAGGCGCCACATCGCGCG 1197  
DB 380 YsGlyProArgArgIleValYsCysPheAsnYsGlyYsGluGlyYrHleIleAlaArg 399  
QY 1198 AACTCGCGCGCCCGCGCAAGAGGCTGCTGAAGTGGCGCAAGAGGCGCCAGCATG 1257  
DB 400 AsnYsArgAlaProArgYsGlyYsGlyStrpYsCysGlyYsGluGlyYrHleIleMet 419  
QY 1258 AAGACTGACCGAGCGCGCCCACTTCTGGGCAAGATCTGCGCCAGCCAGCAAGGCG 1317  
DB 420 YsAspCysThrGluArgGlnAlaAsnPheLeuGlyYsIleTrpProSerHleYsGly 439  
QY 1318 CGCGCCCGGCAATTTCTCTGAGAGCGCGCCGCGCAGCCCGCGCGCGAGAGCTTC 1377  
DB 440 ArgProGlyAsnPheLeuGlnAsnAsnProGluProThrAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTGAAGAGACCAACCCCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
DB 460 ArgPheGluGluThrThrProAlaProYsGlnGluProIleGluArgGluProLeuThr 479  
QY 1438 AGCTGAAGAGCTGTTGGCAACGACCCCTGAGCCAG 1476  
DB 480 SerLeuYsSerLeuPheGlySerAspProLeuSerGln 492  
RESULT 3  
ADP02848  
ID ADP02848 standard; protein; 492 AA.  
XX  
XX ADP02848;  
AC  
XX  
DT 18-NOV-2004 (first entry)

XX Wild type HIV-1 gag protein from strain Du499.  
DE anti-HIV, vaccine, HIV, Gag, prophylaxis, immunogenic response.  
KM Human immunodeficiency virus 1.  
XX WO2004050691-A2.  
XX 17-JUN-2004.  
XX 04-DEC-2003; 2003WO-IB005634.  
XX 04-DEC-2002; 2002ZA-00009830.  
XX (UYCA-) UNITV CAPE TOWN.  
XX (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.  
XX Jaffray A, Williamson A, Rybicki EP;  
XX WPI; 2004-450716/42.  
XX N-PDB; ADP02846.  
XX New vector including a nucleotide sequence encoding the gag polypeptide,  
XX useful as a vaccine for the treatment or prophylaxis of HIV infection in  
XX a mammal.  
XX Disclosure; SEQ ID NO 4; 37pp; English.  
XX The invention relates to a vector including a nucleotide sequence  
XX encoding an HIV gag polypeptide, where the nucleotide sequence encoding  
XX the gag polypeptide comprises a sequence having at least 90% sequence  
XX identity to 1549 (S1) or 1479 (S2) bp defined in the specification. The  
XX vector is useful as a vaccine for the treatment or prophylaxis of HIV  
XX infection in a mammal, where the vaccine induces an immunogenic response  
XX to the virus-like particles in a suitable susceptible host. This sequence  
XX corresponds to an HIV-1 gag protein sequence used in the method of the  
XX invention.  
XX SQ Sequence 492 AA;  
XX  
XX Alignment Scores:  
XX Pred. No.: 2,76e-137 Length: 492  
XX Score: 2441.00 Matches: 459  
XX Percent Similarity: 97.16% Conservative: 20  
XX Best Local Similarity: 93.10% Mismatches: 12  
XX Query Match: 88.06% Indels: 2  
XX DB: Gaps: 2  
XX  
XX US-09-475-704A-3 (1-1479) x ADP02848 (1-492)  
QY 1 ATGGGCGCGCGCGCATCTGCGCGCGCGGAGAGCTGAGACCGCTGGAGCGCATCCGCGC 60  
DB 1 MetGlyAlaArgPheSerLeuMetGlyGlyIleuAspLysLeuAspLysTrpGluLysIleArg 20  
QY 61 CTGGCGCGCGCGCGGAGAGAGTGTACATGATGAAAGACCTGATGGCGCGCGAG 120  
DB 21 LeuArgProGlyGlyIleuLysHisTrpMetLeuLysHisIleValTrpAlaSerArgLysGlu 40  
QY 121 CTGGAGAGAGTTCGCGCTGAAACCGCGCTGCTGGAGACCGAGAGGCGCTGGCAAGATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuGluTrpSerGluGlyCysLysGlnIle 60  
QY 181 ATCCGCGAGCTGACACCGCGCTGCGAGACCGGACGAGAGAGCTGAAGAGCTGTTCAC 240  
DB 61 MetLysGlnLeuGlnProAlaLeuGlnTrpGlyTrpGluIleuLysSerLeuTrpAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCTGTGACGAGAGAGATGAGGTCCGAGACCAAGAGAGCC 300  
DB 81 ThrValAlaThrLeuTrpCysValHisGluLysIleGluValArgAspThrLysGlnIle 100  
QY 301 CTGGACAGAGTGTGAGAGAGAGAGAGAGAGTCCACGAGAGATCCAGAGGCGAGGCC 360

DB 101 LeuAspLysIleGluGluGlnIleAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 120  
QY 361 GCCGACAAAGGAGGTGAGCCGAACCTAACCCCATGTGCGAAGACCTGACGGCCAGATG 420  
DB 121 AlaAsp---GlyLysValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlyGlnMet 139  
QY 421 GTGGACCAAGGCGCATCAAGCCCGCGCACCTGAAACCGCTGGGTGAAGTATGAGAGAG 480  
DB 140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGluLys 159  
QY 481 GCCTTCAGCCCGGAGGTATCCCATGTTCAACCGGCTGAGCGAGGCGGCCACCCCGAG 540  
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProIle 179  
QY 541 GACCTGAACAGCATTTGAACACCGTGGCGGCGCACAGCGCGCCATCAGATGCTGAAG 600  
DB 180 AspLeuAsnTrpMetLeuAsnThrValGlyGlnHisGlnAlaIleMetGlnMetLeuLys 199  
QY 601 GACACCATCAAGAGAGGCGCGGAGTGGAGACCGCGTGCACCCCGTGCAGCGCGCCC 660  
DB 200 AspTrpIleAsnGluGlnAlaAlaGluTrpAspArgLeuHisProValHisAlaGlyPro 219  
QY 661 ATCGCCCGCGCGAGATGCGGAGCGCGCGCGAGCGGACGACATGCGCGGACCGACGACC 720  
DB 220 IleAlaProGlyGlnMetArgGluProArgLysSerAspIleAlaGlyThrThrSerThr 239  
QY 721 CTGGAGAGAGATGCTGCTGATGACGACGACACCCCGCATCCCGTGGCGGACATTTAC 780  
DB 240 LeuGlnGluGlnIleAlaTrpMetThrSerAsnProIleProValHisAspIleLys 259  
QY 781 AAGCGGTGATCATCTGCGGCTGAAACAAAGATCGGCGAGATGACAGCGCGGTAGCATC 840  
DB 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIle 279  
QY 841 CTGACATCAAGCAAGGCGCGCGAGAGCGCTTCGCGCATGAGACCGCTTCTTCAAG 900  
DB 280 LeuAspIleArgGlnGlyProLysGluProPheArgAspTrpValAspArgPheLys 299  
QY 901 ACCGTGGCGCGCGAGAGAGACCGAGAGGTTGAAGATGAGACCGACACCTTCTG 960  
DB 300 ThrLeuArgAlaGluGlnAlaThrGlnIleValLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGGAGAAAGCGCAACCGCGATGCAAGACCATCTGCGCGCTCGGCGCGCGCGCGCG 1020  
DB 320 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaThr 339  
QY 1021 CTGAGAGATGATGACCGCTGCGAGGCGGTGGCGCGCGCGCGCGCGCGCGCG 1080  
DB 340 LeuGluGluMetCetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgVal 359  
QY 1081 CTGGCGCGAGGAGTGAAGCGGCGCAACAC---AGCGTGAATGATGAGAGAGCACTTC 1137  
DB 360 LeuAlaGluAlaMetSerGlnThrAsnSerGlyAsnIleMetCetGlnArgSerAsnPhe 379  
QY 1138 AAGGCGCGCGCGCATCTGTCAGATGCTTCAACTGCGCGGCAAGAGGCGCATGCGCGC 1197  
DB 380 LysGlyProArgArgIleValLysCysPheAsnLysGlyLysGluGlnHisIleAlaArg 399  
QY 1198 AACTGCGCGCGCGCGCGAGAGAGGCTGCTGAAGTGGCGCAAGAGGCGCACAGATG 1257  
DB 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlnHisGlnMet 419  
QY 1258 AAGACTGACCGGAGCGCGCGCAACTTCCTGGGCAAGATCTGGCGCGCGCGCGCGCG 1317  
DB 420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGly 439  
QY 1318 CGCCCGCGCAACTTCTGAGAGAGCGCGCGCGAGCGCACCGCGCGCGCGAGAGCTTC 1377  
DB 440 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGluSerPhe 459  
QY 1378 CGCTTCAGAGAGACACCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
DB 460 ArgPheGluGlnThrThrProAlaProLysGlnGluProIleGluArgLysProLeuThr 479



QY 1258 AAGGACTGCACGAGCGCCAGGCAACTTCTGGGGCAAGATCTGGCCCAAGGCAAGGGC 1317  
DB 422 LysAspGlyThrGluThrGlnAlaAsnGluGlyLysIleTyrProSerHisLysGly 441  
QY 1318 CGCCCGGCAACTTCTGCAGAGCGCCCGGCAAGCCAGCCCGCCCGCGAGAGCTTC 1377  
DB 442 ArgProGlyAsnGluGluGlnAsnArgProGluProThrAlaProProAlaGluSerPhe 461  
QY 1378 CGCTTGAAGAGACCAACCCCGGCGCAAGACAGAGAGACGAGAGCCGAGACCTGACC 1437  
DB 462 ArgPheGluGluThrThrProAlaProLysGlnGluProIleGluArgGluProLeuThr 481  
QY 1438 AGCCTGAAGAGCCTGTTCCGCAAGACCCCTGAGCCAG 1476  
DB 482 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 494  
RESULT 5  
ADN36396  
ID ADN36396 standard; protein; 514 AA.  
XX  
AC ADN36396;  
XX  
DT 15-JUL-2004 (first entry)  
XX  
DE Human protein for anti-HIV vaccine.  
XX  
KM anti-HIV, vaccine; HIV; promoter; viral particle; immunization.  
XX  
OS Homo sapiens.  
XX  
PN WO2004035006-A2.  
XX  
PD 29-APR-2004.  
XX  
PF 17-OCT-2003; 2003WO-US03112.  
XX  
PR 18-OCT-2002; 2002US-0419465P.  
XX  
PA (AARO-) AARON DIAMOND AIDS RES CENT.  
XX  
PI Huang Y, Ho DD, Chen Z;  
DR WPI; 2004-348328/32.  
DR N-PSDB; ADN36395.  
XX  
PT Nucleic acid vector comprising at least one HIV sequence operably linked  
PT to a promoter and encoding a protein that does not assemble into viral  
PT particles, useful in immunizing a subject against HIV infection.  
XX  
PS Disclosure; SEQ ID NO 10; 166p; English.  
XX  
CC The invention relates to a nucleic acid vector comprising at least one  
CC HIV sequence operably linked to a promoter and encoding a protein that  
CC does not assemble into viral particles. The nucleic acid vector is useful  
CC in immunizing a subject against HIV infection. This sequence corresponds  
CC to a peptide used in the invention.  
XX  
SQ Sequence 514 AA;  
Alignment Scores:  
Pred. No.: 5,14e-137 Length: 514  
Score: 2436.50 Matches: 458  
Percent Similarity: 96.95% Conservative: 19  
Best Local Similarity: 93.09% Mismatches: 14  
Query Match: 87.90% Indels: 1  
DB: 8 Gaps: 1  
US-09-475-704A-3 (1-1479) x ADN36396 (1-514)  
QY 1 ATGGGCGCGCGCGCAGAGATCTTGCGCGCGCGCAGCTGAGACGCTTGAGAGCGCATCCGC 60  
DB 24 MetGlyAlaArgAlaSerIleuAlaGlyGlyLysLeuAspLysTyrGluLysIleArg 43

QY 61 CTGCGCGCGCGCGCGCAAGAGTCTTACATGATGAAGCACTGTGTGTGGCCAGCCCGCAG 120  
DB 44 LeuArgProGlyGlyLysLysHisTyrMetLeuLysHisLeuValTyrPheSerArgGlu 63  
QY 121 CTGAGAAAGTTGGCCTGAACCCCGGCGTGGAGACAGAGAGGGGTGCAAGAGATC 180  
DB 64 LeuGluArgPheAlaLeuAsnProGlyLeuGluGlnTrpSerGluGlyCysLysGlnIle 83  
QY 181 ATCCGCGCAGCTGACCCCGCGCTGAGACCGGCGACGAGAGAGCTGAAGAGCCTGTTCAAC 240  
DB 84 IleLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluGluLeuArgSerLeuPheAsn 103  
QY 241 ACCGCGCGCAGCCTTACTATGCTGTGACAGAGAAAGATCGAGGTCCCGGACACCAAGAGGCC 300  
DB 104 ThrValAlaThrLeuTyrCysValHisGluGlyIleGluIleArgAspThrLysGluAla 123  
QY 301 CTGACAAGATCGAGAGAGAGACAGAAAGTGGCCAGAGATCCAGAGATCCAGAGCGGAGGCC 360  
DB 124 LeuAspLysIleGluGluGlnHisnLysIleGlnGlnLysThrGlnGlnAlaLysLys 143  
QY 361 GCCGCAAGAGCGCAAGTGAAGCCCAAGACTACCCCATCTGACAGAACTTGACAGGCGCAGATG 420  
DB 144 AlaAspGlu--LysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMet 162  
QY 421 GTGACCGAGCGCATCAGCGCCCGGCAAGCCCTGAAGCCCTGGGTGAGAGTGTATGAGAGAG 480  
DB 163 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValIleGluGlnLys 182  
QY 481 GCCTTCAGCCCGCGAGTGATCCCATGTTCACCGCCCTGAGCGAGGCGGCCACCCCGCAG 540  
DB 183 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlyAlaThrProGln 202  
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGGCCCAAGCGCGCCATGACAGATCTGAAG 600  
DB 203 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 222  
QY 601 GACACCATCAAGAGAGAGCGCGCGAGTGGAGACCGCGTGCACCCCGTGCACCGCGGCGCC 660  
DB 223 AspThrIleAsnGluGluAlaAlaGluTyrAspArgValHisProValHisAlaGlyPro 242  
QY 661 ATCGCCCGCGCGCAGATGCGGAGCGCCCGCGCGCAGCATCGCGCGCACCCAGCAC 720  
DB 243 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 262  
QY 721 CTGCGAGAGCAGATCGCTGTGATGACAGCAACCCCGCATCCCGTGGCGGCGCATCTAC 780  
DB 263 LeuGlnGlyGlnIleAlaTyrMetThrSerAsnProProValProValGlyGluIleTyr 282  
QY 781 AAGCGGTGATCATCTGGGCGCTGACACAGATGTTGCGATGTACAGCCCGTGAAGATC 840  
DB 283 LysArgTyrPheIleIleGlyLysAsnLysIleValArgMetCysSerProValSerIle 302  
QY 841 CTGACATCAAGCAGGCGCCCGAGAGCCCTTCGCGCATACGTGACCGCTTCTTCAG 900  
DB 303 LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 322  
QY 901 ACCCTGGCGCGCGCGCAGAGACCCAGAGAGTGAAGAACTGATGACCGACACCCCTGCTG 960  
DB 323 ThrLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTyrMetThrAspThrLeuLeu 342  
QY 961 GTGCAAGAGCGCAACCCCGCATGCGCAAGCATCTGCGGCTCTCGGCGCGCGCGCGCAGC 1020  
DB 343 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 362  
QY 1021 CTGAGAGAGATGATGACCGCTGCGAGAGCGGTGGCGCGCCAGCACAAGCGCCGCTG 1080  
DB 363 LeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHisLysValArgVal 382  
QY 1081 CTGGCCAGAGCGCATGAGCCAGGCGCAACACAGCGTATATGCAAGAGACCACTTCAAG 1140  
DB 383 LeuAlaGluAlaMetSerGlnAlaAsnGlyThrIleLeuMetGlnArgSerAsnPheLys 402



```
QY 841 CTGGAGATCAAGAGGCCCCCAAGAGCCCTTCCGCGACTACGTGAGCCGCTTCTTGAG 900
DB 280 LeuAspIleArgGlnGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 299
QY 901 ACCCTCGCGCCGAGAGAGACACCCAGAGGTAAAGATGATGACCGACACCTGCTG 960
DB 300 ThrLeuArgAlaIleGlnGlnAlaThrGlnGlnValLysAsnTyrMetThrAspThrLeuLeu 319
QY 961 GTGCAAGAGCCCAACCCCGACTGCAAGACCATCTCGCGGCTCTCGAGCCCGGCGCAGC 1020
DB 320 ValGlnAsnAlaAsnProAspCysLysTyrThrIleLeuArgAlaLeuGlnProGlnAlaThr 339
QY 1021 CTGGAGAGATGATGACCCGCTGCGAGGCGTGGGCGGCGCCGAGCCCAAGGCGCCGCTG 1080
DB 340 LeuGlnGlnMetThrAlaCysGlnGlnValGlyGlyProGlnHisLysAlaArgVal 359
QY 1081 CTGGCCGAGCGCATGAGCCAGCCACACCC---AGCGTATGATGAGAGAGCAATTC 1137
DB 360 LeuAlaGlnAlaMetSerGlnThrAsnSerGlyAsnIleMetSerGlnArgSerAsnPhe 379
QY 1138 AAGGCGCCCGCGGCGCATCTGTCAGAGTCTTCATCTGCGGCAAGAGGCGCAATCGCCGC 1197
DB 380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGlnGlnHisIleAlaArg 399
QY 1198 AACTGCGCGCGCCCGCGCAAGAGGCGTGTGAGAGTGGGCAAGAGGCGCCACGAGATG 1257
DB 400 AsnCysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGlnGlnHisGlnMet 419
QY 1258 AAGAGTCGACCGAGCGCCAGGCGCAATCTCTGGGCAAGATCTGGCCCAAGAGGCGC 1317
DB 420 LysAspCysThrGlnArgGlnAlaAsnPheLeuGlnLysIleTyrProSerHisLysGly 439
QY 1318 CGCCCGCGCAACTTCTTGAGAGCCCGCCGAGCCGAGCCGCGCCCGCCGAGAGCTTC 1377
DB 440 ArgProGlnLysAsnPheLeuGlnAsnArgProGlnProThrAlaProProAlaGlnSerPhe 459
QY 1378 CGCTTGAGAGAGACACACCCCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1437
DB 460 ArgPheGlnGlnThrThrProAlaProLysGlnGlnProIleGlnArgGlnProLeuThr 479
QY 1438 AGCTGAGAGAGCTGTGTGCGCAACGACCCCTGAGGCGAG 1476
DB 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492
RESULT 7
AAM48948
ID AAM48948 standard; protein; 635 AA.
XX
AC AAM48948;
XX
DT 29-AUG-2003 (revised)
DT 19-APR-2002 (first entry)
XX
DE HIV-1 subtype C isolate Du422 gag protein.
XX
KW HIV-1 subtype C; vaccine; HIV infection; AIDS; gag; antiviral.
XX
OS Human immunodeficiency virus; type I.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 523 /note= "encoded by TAA"
XX FT Misc-difference 526 /note= "encoded by TAG"
XX FT Misc-difference 557 /note= "encoded by TAA"
XX FT Misc-difference 612 /note= "encoded by TGA"
XX FT Misc-difference 612 /note= "encoded by TGA"
XX PN MO20020494-A2.
XX PD 17-JAN-2002.
```

```
XX
PF 09-JUL-2001; 2001MO-1B001208.
XX
PR 07-JUL-2000; 2000US-0216995P.
XX
PR 10-JUL-2000; 2000ZA-00003437.
XX
PR 15-SEP-2000; 2000ZA-00004924.
XX
PA (MEDICAL RES COUNCIL.
XX (UICA-) UNIT CAPE TOWN.
XX (UYN-) UNIT NORTH CAROLINA.
XX
PI Williamson C, Swanstrom RJ, Morris L, Karim SA, Johnston RE;
XX
DR WPI: 2002-171700/22.
XX
DR N-PSDB; AAL41591.
XX
XX
PT Selecting HIV-1 subtype C isolates, which are useful in developing
PT vaccines against HIV infection, comprises isolating viruses with high
PT sequence identity to a consensus sequence whose phenotype is associated
PT with the HIV subtype.
XX
XX
PS Claim 17; Page 64-66; 69p; English.
XX
XX
CC The present invention relates to a process for the selection of human
CC immunodeficiency virus (HIV) subtype isolates for use in the development
CC of a prophylactic and/or therapeutic pharmaceutical composition. The
CC process involves selecting isolated virus or viruses with a high sequence
CC identity to a consensus sequence and a phenotype which is associated with
CC transmission for the particular HIV subtype. The composition can be used
CC in the production of vaccines against HIV. The present sequence is the
CC HIV-1 subtype C isolate Du422 gag protein. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
SQ Sequence 635 AA;
XX
Alignment Scores:
Pred. No.: 6,48e-137 Length: 635
Score: 2435.00 Matches: 458
Percent Similarity: 96.96% Conservative: 20
Best Local Similarity: 92.90% Mismatches: 13
Query Match: 87.84% Indels: 2
DB: 5 Gaps: 2
US-09-475-704A-3 (1-1479) x AAM48948 (1-635)
QY 1 ATGGGCGCCCGCGCGCATCTCTGCGCGCGCGCAAGCTGAGCGCTGAGAGCGCATCCGC 60
DB 31 MetAlaAlaArgAlaSerIleLeuArgGlyGlnLysLeuAspLysTyrGlnLysIleArg 50
QY 61 CTGGCGCCCGCGCGCAAGAGTGTACATGATGAAGACCTGTGTGGCGCCAGCGCGAG 120
DB 51 LeuArgProGlnGlyLysLysHisTyrMetLeuLysHisIleValThrAlaSerArgGln 70
QY 121 CTGAGAGATTGCGCTGAAACCCCGCGCTGTGAGAGACGAGCGCTGAGACGATC 180
DB 71 LeuGlnArgPheAlaLeuAsnProGlnLeuLeuGlnThrSerGlnGlyCysLysGlnIle 90
QY 181 ATCCGCGAGCTGCAACCCCGCGCTGCAAGCGCGGAGAGAGAGTGAAGAGCTTCAAC 240
DB 91 MetLysGlnLeuGlnProAlaLeuGlnThrGlnThrGlnLysSerLeuTyrAsn 110
QY 241 ACCGCGCGCACCTGTACTGCTGTCACAGAAAGATCGAGCTCCGCGACACCAAGAGAGCC 300
DB 111 ThrValAlaThrLeuTyrCysValHisGlnLysIleGlnValAlaGAspThrLysGlnAla 130
QY 301 CTGACAAAGATCGAGAGAGACCAAGACAGTGTGAGAGAGATTCAGAGCGCGAGGCC 360
DB 131 LeuAspLysIleGlnGlnGlnGlnAsnLysCysGlnGlnLysThrGlnGlnAlaLysAla 150
QY 361 GCCGCAAGGCGCAAGTGTGAGAGACCAAGTCAACCCCATCTGTGAGAGAGCTGAGGCGCGATG 420
DB 151 AlaAsp---GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 169
```

```
QY 421 GTGCACAGGCGCATCAGCCCCCGCACCTGTAAAGCGCTGAGTGAAGTATCGAGAGAAAG 480
DB 170 ValHisGlnIleAlaIleSerProArgThrLeuAsnAlaTrrValIleGlnGluLys 189
QY 481 GCGTTGAGCCCGGAGGTGATCCCATGTTTCAAGCCCGTGAAGGAGGCGCACCCCGAG 540
DB 190 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGlnGluAlaThrProGln 209
QY 541 GACCTGAACACGATGTTGAAACACCGTGGGGGGGCGCACAGGCGCCCATGCAAGATCGTGAAG 600
DB 210 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaIleMetGlnMetLeuLys 229
QY 601 GACACCATCAACAGAGAGGCGCGGAGTGGAGCCGGGTGACCCCGTGCAGCGCGCGCC 660
DB 230 AspThrIleAsnGlnIleAlaIleGlnIleTrrPheArgLeuHisProValHisAlaGlyPro 249
QY 661 ATGCGCCCGGCGGAGTGGCGGAGCGCGGCGGAGGAGCATGCGCGGAGCACACGAGCACC 720
DB 250 IleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThr 269
QY 721 CTGCGAGGAGGAGGATGCGCTGAGTGAAGCAACCCCGCATCCCGCTGGGCGGAGCATTCAG 780
DB 270 LeuGlnGlnGlnIleAlaTrrMetTrrSerAsnProProIleProValGlyAspIleTyr 289
QY 781 AAGCGGTGATCATCTCGGCGCTGAACAAGATGTCGCGATGATGACAGCCCGTGAAGCATC 840
DB 290 LysArgTrrIleIleLeuGlyLeuAsnLysIleValArgMetTrrSerProValSerIle 309
QY 841 CTGAGCATCAAGAGGCGCGCGCAAGAGCCCTTCGCGACTACGTGAGCCGCTTCTTCAAG 900
DB 310 LeuAspIleArgGlnGlyProLysGluProPheArgAspTrrValAspArgPhePheLys 329
QY 901 ACCCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
DB 330 ThrLeuArgAlaGlnIleAlaThrGlnGlnValLysAsnTrrPheThrAspThrLeuLys 349
QY 961 GTGCAGAAAGCGCAACCCCGACTGCAAGACATCTCGCGCTTCGCGCGCGCGCGCGCAGC 1020
DB 350 ValGlnAsnAlaAsnProAspGlySerThrIleLeuArgAlaLeuGlyProGlyAlaThr 369
QY 1021 CTGAGAGAGATGATGACCCCGCTGCGAGGCGTGGCGCGCGCGCGCGCGCGCGCGT 1080
DB 370 LeuGlnGlnMetMetThrAlaCysGlnGlyValGlyProGlyHisLysAlaArgVal 389
QY 1081 CTGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1137
DB 390 LeuAlaGlnAlaMetSerGlnThrAsnSerGlyAsnIleMetGlnTrrSerAsnPhe 409
QY 1138 AAGGCGCGCGGCGGAGTGTCAAGTCTTCAACTGCGGAGGAGGAGGAGGAGGAGGAGGAG 1197
DB 410 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGlnGlyHisIleAlaArg 429
QY 1198 AACTGCGCGCGCGCGCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1257
DB 430 AsnCysArgAlaProArgLysLysGlyCysTrrPheCysGlyLysGlnGlyHisGlnMet 449
QY 1258 AAGGAGTGAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1317
DB 450 LysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrrProSerHisLysGly 469
QY 1318 CCGCGCGGAGCACTTCTGAGAGAGCGCGCGGAGCGGAGCGCGCGCGCGCGGAGCTTC 1377
DB 470 ArgProGlyLysAsnPheLeuGlnAsnArgProGlnProThrAlaProProAlaGlnSerPhe 489
QY 1378 GCGTTGAGAGAGCAACCCCGCGCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1437
DB 490 ArgPheGlnGlnIleThrThrProAlaProLysGlnGlnProIleGlnLysGlnProLeuThr 509
QY 1438 AAGCTGAAGAGGCTGTTGCGCAACGAGCCCGCTGAGGAGCAG 1476
DB 510 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 522
RESULT 8
```

```
ADCT2880
ID ADCT2880 standard; protein; 492 AA.
XX
AC ADCT2880;
XX
DT 18-DEC-2003 (first entry)
XX
DE HIV-1 gag protein containing an NCP sequence, SEQ ID NO 110.
XX
KW complex; HIV nucleocapsid protein 7; NCP7; HIV-ps1-site; anti-HIV;
KW human immunodeficiency virus; HIV.
XX
OS Human immunodeficiency virus 1.
XX
PN MO2003060098-A2.
XX
PD 24-JUL-2003.
XX
PF 10-JAN-2003; 2003WO-US000801.
XX
PR 11-JAN-2002; 2002US-0347369P.
XX
PA (AChI-) ACHILLION PHARM INC.
XX
PI Beutler D, Hou X, Marlor CW, Rice WG, Yang W;
XX
DR WPI; 2003-646042/61.
XX
PS Claim 2; SEQ ID NO 110; 105pp; English.
XX
CC The invention relates to a novel method of determining whether a compound
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)
CC polypeptide and an HIV-ps1-site oligonucleotide. The method involves
CC adding the HIV-ps1-site oligo to a mixture of the NCP7 polypeptide and
CC the novel compound and comparing amount of complex formed, with that
CC formed in the absence of the compound, where a decrease in the amount of
CC complex formed in presence of the compound indicates that the compound
CC inhibits complex formation. The method of the invention involves anti-HIV
CC activity. The compounds identified are useful for treating a subject
CC infected with human immunodeficiency virus (HIV) by administering the
CC compound to the subject. This sequence represents an HIV-1 gag protein
CC which contains an NCP sequence of the invention.
XX
SQ Sequence 492 AA;
XX
Alignment Scores:
Pred. No.: 3,26e-136
Score: 2423.00 Length: 492
Percent Similarity: 96.96% Matches: 451
Best Local Similarity: 87.41% Mismatches: 27
Query Match: 97.41% Indels: 13
DB: Gaps: 2
US-09-475-704A-3 (1-1479) x ADCT2880 (1-492)
QY 1 ATGCGCGCGCGGCGGAGTCTGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspThrTrrGlnLysIleArg 20
QY 61 CTGCGCGCGCGGCGGAGAGTGTCAAGTGAAGAGCACTGCTGGAGGAGGAGGAGGAGGAGGAG 120
DB 21 LeuArgProGlyGlyLysLysCysTrrMetLeuLysHisIleValTrrAlaSerArgGln 40
QY 121 CTGAGGAAGTTCGCGCTGAACCCCGGCTGCTGAGAGAGCAAGGAGGAGGAGGAGGAGGAG 180
DB 41 LeuGlnArgPheSerLeuAsnProGlyLeuGlnGlnTrrSerGlnGlyCysLysGlnIle 60
QY 181 ATCGCGAGCTGCAACCCCGCTGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
```



```

Db      61 MetValGlnLeuGlnProAlaLeuGlnThrGlnGluLeuLysSerLeuTyrasn 80
      241 ACCGTGGCCACCTGTACTGCGTGCAGAGAGATGAGGTCCGACACCAAGAGGCC 300
      81 ThrValAlaThrLeuPheCysValHisGluLysIleAlaValArgAspThrLysGluAla 100
Qy      301 CTGGACAAGATCGAGAGAGAGAGAACAGTGCACACAGATCCAGAGCGCCGAGGCC 360
      101 LeuAspLysIleGlnGluGlnGlnAsnLysSerGlnGlnLysThrGlnGlnAlaLysAla 120
Qy      361 GCCGACAAGGGGCAAGGTGAGCCAGAACTACCCCATGTGTCAGAACTGAGAGGCCGATG 420
      121 AlaAsp---GlyThrValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet 139
Qy      421 GTGCACACGAGCCATCGACCCCGCCACCTGACGCGCTGGGTGAAGGTGATCGAGAGAG 480
      140 ValHisGlnAlaIleSerProArgThrLeuAsnAlaTyrValLysValIleGlnGluLys 159
Qy      481 GCCTTAGCCCGGAGGTGATCCCATGTTCAACCGCCCTAGAGAGAGGCCACCCCGCCAG 540
      160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGlnGluAlaThrProGln 179
Qy      541 GACCTGAACACAGATGTTGAACAACCTGGGCGGCGCCACAGCGCCATGCGAGATGCTGAAG 600
      180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 199
Qy      601 GACACCATCAAGAGAGAGGCCCGCCAGTGGAGCCGCGTGCACCCCGTGCACCGCGGCGCC 660
      200 AspThrIleAsnGlnGlnIleAlaGlnTyrAspArgLeuHisProValHisIleGlyPro 219
Qy      661 ATGCGCCCGGCGCAGATGCGCGAGCCCGCGGAGAGACATGCGCGGCGCCACCGACCG 720
      220 IleAlaProGlyGlnMetArgLysProArgLysGlySerAspIleAlaGlyThrThrSerThr 239
Qy      721 CTGCAGAGACAGATGCGCTGATGATGACCAACACCCCGCCATCGCGTGGCGCACTTAC 780
      240 LeuGlnGlnGlnIleAlaTyrMetThrAsnAsnProProValProValGlyAspIleTyr 259
Qy      781 AAGCGGTGATGATCTGTGGGCTGAAACAAGATCTGCGGATGATACACCCCGTGAAGCATC 840
      260 LysArgTyrPheIleLeuGlnLysLeuAsnLysIleValArgMetLysSerProValSerIle 279
Qy      841 CTGCACATCAAGAGAGAGGCCCGCCAGAGCCCTCCGCACTAAGTGAACCGCTTCTTCAAG 900
      280 LeuAspIleArgGlnGlyProLysGlnProPheArgAspTyrValAspArgPhePheLys 299
Qy      901 ACCCTGCGCGCGGAGAGAGCAACCCAGAGGTGAAGACTGATGATCCGACACCTGCTGT 960
      300 ThrLeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTyrMetThrAspThrLeuLeu 319
Qy      961 GTGCAGAAAGCCCAACCCCGCATGCAAGACATCTGCGGCGCTGCGGCGCGCGCCAGC 1020
      320 ValGlnAsnAlaAsnProPheCysLysThrIleLeuAspAlaLeuGlnProGlyAlaThr 339
Qy      1021 CTGAGAGAGATGATGACCGCTGCAAGGCGGTGGGCGGCGCCAGCCACAGAGCCGCGTG 1080
      340 LeuGlnGlnMetMetThrAlaCysGlnGlyValGlyLysProSerHisLysHisAlaArgVal 359
Qy      1081 CTGGCGGAGGCGATGAGCCAGGCC---AAGACCAAGCTGATGATGCAAGAGCAACTTC 1137
      360 LeuAlaGlnLysLysSerGlnThrAsnAlaAsnIleMetMetGlnArgSerAsnPhe 379
Qy      1138 AAGGCGCCCGCGCATGCTCAAGTGTCTCACTGCGGCAAGAGAGGCGCCATGCGCCGC 1197
      380 LysGlyProArgArgIleIleLysCysPheAsnCysGlyLysGlnGlyHisLeuAlaArg 399
Qy      1198 AATCCCGCGCGCCCGCAAGAGAGGCTGCTGAAGTGGCGGCAAGAGGCGCCACAGATG 1257
      400 AsnCysArgAlaProArgLysLysGlyCysTyrLysCysGlyLysGlnGlyHisGlnMet 419
Qy      1258 AAGGACTGACCGGACCGCCAGGCCCACTTCTGGGCAAGATGGCCCAAGCCACAGGCG 1317

```

```

Db      420 LysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTyrProSerHisLysGly 439
Qy      1318 CGCCCGGCAACTTCTCTGCAGAGCCGCCGAGCCACCGCCCGCCCGCGAGAGCTTC 1377
      440 ArgProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProAlaGlnSerPhe 459
Qy      1378 CGCTTCGAGAGAGACACACCCCGCCAGAGAGAGACACAGACCGCGACACCTGACC 1437
      460 ArgPheGlnGluThrThrProAlaProLysGlnGlnProArgGluArgGluProLeuThr 479
Qy      1438 AGCTGAAGAGCTGTTCGCAACGACCCCTGAGCCAG 1476
      480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492
Db
RESULT 9
ADN36406
ID      ADN36406 standard, protein, 1457 AA.
AC      ADN36406;
DT      15-JUL-2004 (first entry)
DE      Human protein for anti-HIV vaccine.
KW      anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
OS      Homo sapiens.
PN      WO2004035006-A2.
PD      29-APR-2004.
PE      17-OCT-2003; 2003WO-US033112.
PR      18-OCT-2002; 2002US-0419465P.
PA      (AARO-) AARON DIAMOND AIDS RES CENT.
XX      Huang Y, Ho DD, Chen Z;
XX      WPI; 2004-348328/32.
XX      DR      N-PSDB; ADN36405.
XX      PT      Nucleic acid vector comprising at least one HIV sequence operably linked
XX      to a promoter and encoding a protein that does not assemble into viral
XX      particles, useful in immunizing a subject against HIV infection.
XX      PS      Disclosure; SEQ ID NO 20; 166pp; English.
XX      CC      The invention relates to a nucleic acid vector comprising at least one
XX      HIV sequence operably linked to a promoter and encoding a protein that
XX      does not assemble into viral particles. The nucleic acid vector is useful
XX      in immunizing a subject against HIV infection. This sequence corresponds
XX      to a peptide used in the invention.
XX      SQ      Sequence 1457 AA;
Alignment Scores:
Pred. No.:      6e-136      Length:      1457
Score:          2419.50      Matches:      455
Percent Similarity: 96.34%      Conservative: 19
Best Local Similarity: 92.48%      Mismatches: 17
Query Match:    87.28%      Indels:      1
DB:             8          Gaps:      1
US-09-475-704A-3 (1-1479) x ADN36406 (1-1457)
Qy      1      ATGGGCGCCCGCGCCAGCATCTGCGGCGGCAAGCTGAGCGCTGAGAGCGCATCCGC 60
      24 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspLysTyrGluLysIleArg 43
      61 CTGGCGCCCGCGGCAAGAGAGTGTACTGATGAGAGCACTGTGTGGGCGGCGCGGAG 120

```





Pred. No.: 8.52e-136 Length: 492  
 Score: 2416.00 Matches: 455  
 Percent Similarity: 95.74% Conservative: 19  
 Best Local Similarity: 92.29% Mismatches: 17  
 Query Match: 87.16% Indels: 2  
 DB: 7 Gaps: 2

US-09-475-704A-3 (1-1479) x ADC72879 (1-492)

```

QY 1 ATGGGCGCCGCGCCAGCATCTCGCGGCGGCAAGCTGAGCCCTGGAGCCGATCCG 60
DB 1 MetGlyAlaSerAlaSerIleLeuArgGlyGlyLeuSerIleArgGlySerIleArg 20
QY 61 CTGGGCGCCGCGGCAAGAGTGTACATATAGACACCTGTGTGGCCAGCCCGAG 120
DB 21 LeuArgProGlyIleGlyValGlyValGlyLeuValIleValIleValIleVal 40
QY 121 CTGGAGAACTTCCGCTGAAACCCGCGCTGTGAGACGACGAGGAGCTGACAGATC 180
DB 41 LeuGluArgPheAlaLeuSerGlyLeuLeuGluThrAlaGluGlyCysGlySerIle 60
QY 181 ATCCGCGACGTGACACCCGCTGACAGCCGCGACGAGAGCTGAAGCCTGTTCAAC 240
DB 61 IleGlyGluLeuGluProAlaLeuGluThrGlyThrGluGluLeuSerLeuTyrAsn 80
QY 241 ACCGTGACCACTCTGACTGCTGACGAGAGATGAGAGTCCGCGACCAAGAGAGCC 300
DB 81 ThrValAlaThrLeuTyrCysValIleAlaGlyIleGluValAlaArgPheThrIleVal 100
QY 301 CTGGACAAAGATCGAGAGAGAGAGCAAGATGCGACAGAGATCCAGAGCCGAGAGCC 360
DB 101 LeuArgPheIleGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 120
QY 361 GCGGACAAAGGCGAGGTGAGCCGAACTACCCCATGTGCGAAGCCTGCAAGGCGAGATG 420
DB 121 AlaAsp---GlyLeuValSerGlnAsnTyrProIleValGluAsnLeuGluGluMet 139
QY 421 GTGACACGAGGCTGAGCCGCGGACCTGGAACGCTGAGGATGATTCAGAGAGAG 480
DB 140 ValIleGluProIleSerProAlaGluThrLeuAsnAlaThrValIleGluGluVal 159
QY 481 GCGTTGACGCGCGGAGGTGATCCCATGTTACCGCCCTGAGGAGAGGCGCCACCCCGAG 540
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGluAlaThrProGlu 179
QY 541 GACCTGAACAGATGTTGAACACCGTGGGCGGCGACAGCCGCGCATGCGATGCTGAAG 600
DB 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyIleGluAlaIleMetGluMetLeuVal 199
QY 601 GACACCATCAAGAGAGGCGCGGAGTGGAGCCGCGTACACCCGCGTACGCGCGCCG 660
DB 200 AspThrIleAsnGluGluAlaIleGluGluGluGluGluGluGluGluGluGlu 219
QY 661 ATCGGCGCGCGGCGAGTGGCGAGCCGCGGCGAGGAGATGCGCGGCGACCAAGAGCC 720
DB 220 ValAlaProGlyIleMetArgIleProArgGlySerAspIleAlaGlyThrThrSerThr 239
QY 721 CTGGAGAGCAGATGCGCTGATGACGAGCAACCCCGCATCCCGTGGCGGACATCTAC 780
DB 240 LeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 259
QY 781 AAGCGGTGATCATCTGGGCTGGAACAAGATCGTGGGATGTACAGCCCGTGGAGATC 840
DB 260 LysArgTyrIleIleLeuGlyLeuAsnValIleValArgMetTyrSerProValSerIle 279
QY 841 CTGGACATCAAGAGGCGCGGAGAGCCCTTCGCGGACTACGTGAGACCGCTTCTTCAAG 900
DB 280 LeuArgPheIleGlyGlyGlyProGlyProPheArgPheTyrValAlaAspArgPhePhe 299
QY 901 ACCGTGCGGCGCGAGAGAGCAACGAGAGGTGAAGAACTGTGATGACGAGACCGCTGTG 960
DB 300 ValLeuArgAlaGluGluAlaThrGluAsnValIleValSerThrThrAspThrLeuLeu 319
  
```

```

QY 961 GTTGAAGACGCCAACCCCGCATGTGCAAGACCATCTGCGCGCTCTGCGCCGCGGCGCAC 1020
DB 320 ValGluAsnAlaAsnProAspCysLeuThrIleLeuArgAlaLeuGlyProAlaIleSer 339
QY 1021 CTGGAGAGATGATGACCGGCTGCGAGGCGCGGCGGCGCGCCGACCAAGCCCGCTG 1080
DB 340 LeuGluGluMetThrAlaCysGluGluValGlyGlyProSerHisLeuValIleArgVal 359
QY 1081 CTGGCGGAGCGATGAGCGAGCGCAAC--ACAGCGTATGATGATGACAGAGCAACTTC 1137
DB 360 LeuAlaGluAlaMetSerGluAlaAsnThrThrAsnIleMetMetGluLysSerHisPhe 379
QY 1138 AAGGCGCGCGCGCATGTCTCAAGTCTTCAACTCGCGCAAGAGGCGCAATGCGCCGC 1197
DB 380 LysGlyProArgArgThrValIleCysPheAsnGlyGlyGluGluGluIleAlaLys 399
QY 1198 AACTCGCGCGCGCGCGCAAGAGGCGTGAAGTGCAGGCGCAAGAGGCGCAAGAGATG 1257
DB 400 AsnGlyArgAlaProArgLysLeuGlyCysThrLysGlyGlyGluGluGluGluMet 419
QY 1258 AAGACTCACCGAGCGCGCAAGCGCAACTCTCGGCAAGATCTGCGCCAGCGCACAAAGGC 1317
DB 420 LysAspCysThrGluArgGluAlaAsnPheLeuGlyLysIleTyrProSerHisLeuGly 439
QY 1318 GCGCGCGCGCACTTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGAGAGCTTC 1377
DB 440 ArgProGlyAsnPheLeuGluSerArgProGluProThrAlaProProAlaGluSerPhe 459
QY 1378 CGCTTGAAGAGACCAACCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
DB 460 ArgPheGluGluThrThrProAlaProLysGluGluGluGluGluGluGluGluGlu 479
QY 1438 AGCTGAAGAGCTGTGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1496
DB 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492
  
```

RESULT 11  
 AAB69275  
 ID AAB69275 standard; protein, 494 AA.  
 AC AAB69275;  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 DE HIV-1 non-subtype B clone 962M651-8 gag protein.  
 XX  
 KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpr;  
 KW vif; vpr; tat; rev; nef; vaccine.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 FN WO200026416-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 25-OCT-1999; 99MO-US024837.  
 ER 02-NOV-1998; 98US-00184418.  
 XX  
 PA (UABR-) UAB RES FOUNDED.  
 XX  
 PI Hahn BH, Shaw GM, Gao F;  
 XX  
 DR WPI; 2000-365651/31.  
 XX  
 PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
 PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.  
 XX  
 PS Claim 41; Fig 14; 131pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 494 AA:

#### Alignment Scores:

Pred. No.:	8,52e-136	Length:	494
Score:	2416.00	Matches:	455
Percent Similarity:	96.77%	Conservative:	24
Best Local Similarity:	91.92%	Mismatches:	12
Query Match:	87.16%	Indels:	4
DB:	3	Gaps:	3

US-09-475-704a-3 (1-1479) x AAB69275 (1-494)

```

QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGCGCAAGCTGAGCCCTGGAGCGCATCCG 60
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAAspLysTrpGluValArg 20
QY 61 CTGGCGCGCGCGCGCGCAAGAGTCTACATGATGAGACCTGGTGGCGCGCGCGAG 120
DB 21 LeuAAspTrpGlyGlyValArgLysTrpMetIleLeuValLeuValTrpAlaSerArgGlu 40
QY 121 CTGGAGAGAGTGGCGCGCGCGCGCGCGCGCTGCGAGAGAGAGAGAGAGAGAGAGAGATC 180
DB 41 LeuGluArgPheAlaLeuAAspProGlyLeuLeuGluThrSerGluIleValSerGlnIle 60
QY 181 ATCCCGCAAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGluLeuAAspSerLeuTyrAAsn 80
QY 241 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
DB 81 ThrValAlaThrLeuTyrCysValAlaValGlnGluValAlaValArgSerTrpLysGlnAla 100
QY 301 CTGGAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 354
DB 101 LeuAAspArgIleGluGlnGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
QY 355 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414
DB 121 GlnAlaAlaAAsp---GlyValValSerGlnAAsnTyrProIleValGlnAAsnLeuGlnGly 139
QY 415 CAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 474
DB 140 GlnMetValAlaGlnLeuValSerProAlaTrpLeuAAsnAlaTrpValLysValIleGln 159
QY 475 GAGAAGGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534
DB 160 GlnValAlaPheSerProGluValIleProMetPheThrAlaLeuSerGluGlnAlaTrp 179
QY 535 CCCGAGAGCTTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 594
DB 180 ProGlnAAspLeuAAsnThrMetLeuAAsnThrValGlyGlnIleAlaIleMetGlnMet 199
QY 595 CTGAAGAGACACATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 654
DB 200 LeuLysAAspTrpThrIleAAsnGlnGlnAlaIleGlnAlaIleTrpAAspArgLeuAAsnIleValAla 219
QY 655 GCGCCCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 714
DB 220 GlyProIleAlaIleProGlyGlnMetArgGluProArgGlySerAAspIleAlaGlyThrThr 239
QY 715 AGCACCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774
DB 240 SerThrLeuGlnGlnGlnIleAlaIleTrpMetThrSerAAsnProIleProValGlyAAsp 259
QY 775 ATCTACAAGCGGTGATCATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834

```

```

DB 260 IleTyrLysArgTrpIleIleLeuGlyLeuAAsnLysIleValAlaArgMetLysSerProVal 279
QY 835 AGCATCTGAGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 894
DB 280 SerIleLeuAAspIleLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 299
QY 895 TTCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954
DB 300 PheLysTrpLeuArgAlaGlnGlnIleAlaThrGlnGlnValLysAAsnTrpMetThrAAspTrp 319
QY 955 CTGCTGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1014
DB 320 LeuLeuValGlnAAsnAlaAAsnProAAspCysLysTrpIleLeuLysAlaLeuGlyProGly 339
QY 1015 GCGAGCGTGGAGAGAGATGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1074
DB 340 AlaThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 359
QY 1075 CGCGTGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1131
DB 360 ArgValAlaLeuAlaGlnAlaMetSerGlnThrAAsnSerValAAsnIleLeuMetGlnLysSer 379
QY 1132 AACTTCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1191
DB 380 AAsnPheLysGlyAAsnLysArgMetValLysCysPheAAsnCysGlyLysGlnIleHisIle 399
QY 1192 GCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1251
DB 400 AlaArgAAsnCysArgAlaProArgLysGlyGlySerTrpLysCysGlyLysGlnIleHis 419
QY 1252 CAGATGAAGAATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1311
DB 420 GlnMetLysAAspCysThrGluArgGlnAlaAAsnPheLeuGlyLysIleTrpProSerHis 439
QY 1312 AAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1371
DB 440 LysGlyArgProGlyAAsnPheLeuGlnAAsnAAspProGluProThrAlaProProAlaGln 459
QY 1372 AGCTTCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
DB 460 SerPheArgPheGlnGluThrThrProAlaProLysGlnGlnSerLysAAspArgGlnAla 479
QY 1432 CTGACCAAGCTTGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
DB 480 LeuThrSerLeuLysSerLeuAAspPheGlySerAAspProLeuSerGln 494

```

RESULT 12  
 ADCT72910 standard; protein; 491 AA.  
 ADCT72910;  
 18-DEC-2003 (first entry)  
 HIV-1 gag protein containing an NCP sequence, SEQ ID No 140.  
 complex; HIV nucleocapsid protein 7; NCP7; HIV-ps1-site; anti-HIV;  
 human immunodeficiency virus; HIV.  
 OS Human immunodeficiency virus 1.  
 PN WO2003060098-A2.  
 PD 24-JUL-2003.  
 PF 10-JAN-2003; 2003WO-US000801.  
 PR 11-JAN-2002; 2002US-0347369P.  
 PA (ACHI-) ACHILLION PHARM INC.  
 XX Beutcher D, Hou X, Marlor CW, Rice WG, Yang W;



KM human immunodeficiency virus; HIV.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO200306098-A2.  
XX  
PD 24-JUL-2003.  
XX  
PF 10-JUN-2003; 2003WO-US000801.  
XX  
PR 11-JAN-2002; 2002US-0347369P.  
XX  
PA (ACHI-) ACHILLION PHARM INC.  
XX  
PI Beuchter D, Hou X, Marlor CW, Rice WG, Yang W;  
XX  
DR WPI; 2003-646042/61.  
XX  
PT Method of determining whether a compound inhibits formation of complex  
PT between HIV nucleocapsid protein 7 polypeptide and HIV psi-site  
PT oligonucleotide by comparing amount of complex formed in presence/absence  
PT of compound.  
XX  
PS Claim 2; SEQ ID NO 104; 105pp; English.  
XX  
CC The invention relates to a novel method of determining whether a compound  
CC inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)  
CC polypeptide and an HIV-psi-site oligonucleotide. The method involves  
CC adding the HIV-psi-site oligo to a mixture of the NCP7 polypeptide and  
CC the novel compound and comparing amount of complex formed, with that  
CC formed in the absence of the compound, where a decrease in the amount of  
CC complex formed in presence of the compound indicates that the compound  
CC inhibits complex formation. The method of the invention involves anti-HIV  
CC activity. The compounds identified are useful for treating a subject  
CC infected with human immunodeficiency virus (HIV) by administering the  
CC compound to the subject. This sequence represents an HIV-1 gag protein  
CC which contains an NCP sequence of the invention.  
XX  
SQ Sequence 491 AA:  
  
Alignment Scores:  
Pred. No.: 2,73e-135 Length: 491  
Score: 2407.50 Matches: 449  
Percent Similarity: 96.34% Conservative: 25  
Best Local Similarity: 91.26% Mismatches: 17  
Query Match: 86.85% Indels: 1  
DB: 7 Gaps: 1  
  
US-09-475-704A-3 (1-1479) x ADC72874 (1-491)

QY 361 GCCGACAAAGGCGAAGGTGAGCCAGAACTACCCCATGTCGACAACTCGAGGCGCAGATG 420  
DB |||||:::--ValSerGlnMetProLeuValGlnSerGlnGlnGlnMet 139  
QY 421 GTGACACGAGCCATCAGCCCGCAGACCCCTGAGAGCTGTGATGAGAGAA 480  
DB |||||:::--ValSerGlnMetProLeuValGlnSerGlnGlnGlnMet 139  
QY 140 ValHisGlnAlaIleSerProAlaGlnMetAlaTyrValIleGlnGlnGln 159  
QY 481 GCCCTTACGCGCGAGGTGATCCCATGTTACCGCCCTGAGGAGGCGCCACCCCGCAG 540  
DB AlapSerProGlnValIleProMetPheThrAlaLeuSerGlnGlnGlnGlnGlnGln 179  
QY 541 GACCTGAACAGATGTTGAACACCGTGGCGCGCCACAGCCCGCCATGAGATGCTGAAG 600  
DB AapLeuMetThrMetLeuMetThrValGlyGlyHisGlnAlaMetGlnMetLeuLys 199  
QY 601 GACACCATCAAGAGAGGCGCGCGAGTGGAGCCGCGTGCACCCGTCGACCGCGCGCC 660  
DB AapThrIleAsnGlnGlnAlaIleGlnTyrPhePheGlnHisProValHisAlaGlyPro 219  
QY 661 ATCGCCCGCGCGCAAGTGGCGAGCCCGCGGACGACATGCGCGGACCAACGACACC 720  
DB 220 AlaAlaProGlyGlnMetArgLupProArgLysSerAlaGlyThrThrSerThr 239  
QY 721 CTGCAAGAGACATGCGCTGTGATGACACGACACCCCGCCATCCGCTGGCGCCATCTAC 780  
DB 240 LeuGlnGlnGlnIleAlaIleThrMetThrGlyAsnProProValProValGlyAspIleTyr 259  
QY 781 AAGCGGTGATCATCTCGGCGCTGAAACAAGATCGCGGATATACGCGCGCGTGAACATC 840  
DB 260 LysAspGlyPheIleIleLeuGlnLeuMetLysIleValArgMetLysSerProValSerIle 279  
QY 841 CTGACATCAAGCAGGCGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 900  
DB 280 LeuAspIleLysGlnGlyProLysGlnLupPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGCGCGCGCGAGACAGACCCAGAGAGGTGAAGAACTGATGACACGACCTGCTG 960  
DB 300 ValLeuArgAlaGlnGlnAlaThrGlnMetValLysMetThrMetThrMetThrLeu 319  
QY 961 GTGCAAGAGCGCAACCGCGCATGCAAGACATCTCGCGGCTCTGCGCGCGCGCGCAGC 1020  
DB 320 IleGlnMetAlaAsnProAspCysLysThrIleLeuLysAlaLeuGlyProAlaAlaSer 339  
QY 1021 CTGAGAGAGATGATACCGCTGCAAGGCGGTGAGCGCGCGCCGACCAAGGCGCGCTG 1080  
DB 340 LeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgVal 359  
QY 1081 CTGCGCGAGCGCATGAGCCAGGCGCAACACGCGGTGATGATGACAGAAAGCAATTCAAG 1140  
DB 360 LeuAlaGlnAlaMetSerGlnAlaAsnSerAsnIleMetMetGlnArgSerAsnPheLys 379  
QY 1141 GCGCGCGCGCATGATGTAAGTCTTCAACTGCGGCAAGAGGCGCACTGCGCGCGCAC 1200  
DB 380 GlySerLysArgIleValLysCysPheAsnCysGlyLysGlnGlyHisIleAlaArgAsn 399  
QY 1201 TGCCGCGCGCGCGCAAGAGGCTGTGAAAGTGCAGGCAAGAGGCGCCACAGATGAAG 1260  
DB 400 CysArgAlaProArgLysLysGlyCysThrPheLysCysGlyGlnGlnGlyHisGlnMetLys 419  
QY 1261 GACTGACACGAGCGCGCAAGGCGCAACTTCTGGGCAAGATCTGGCGCCAGCAAGGCGCGC 1320  
DB 420 AapCysThrGlnLysArgAlaAlaAsnMetLeuGlyLysIleTyrProSerHisLysGlyArg 439  
QY 1321 CCGCGCAACTTCTGTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
DB 440 ProGlyAsnPheLeuGlnSerThrProGlnProThrAlaProProAlaGlnSerPheArg 459  
QY 1381 TTGAGAGAGACCAACCCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 460 PheGlnGlnThrThrProAlaProLysGlnGlnLysLysAspArgLupProLeuIleSer 479

QY 1441 CTGAAGAGCCCTGTTCCGCAACGACCCCTGAGCCAG 1476  
|||||  
Db 480 LeuLysSerLeuPheGlySerAspProSerSerGln 491  
|||||

## RESULT 14

ID AAB69278 standard; protein; 491 AA.

AC AAB69278;

DT	12-SEP-2003	(revised)
DT	20-APR-2001	(first entry)

DE HIV-1 non-subtype B clone 94IN476-104 gag protein.

KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpr;

OS Human immunodeficiency virus 1.

PN WO200026416-A1

PD 11-MAY-2000.

PF 25-OCT-1999; 99WO-US024837

PR 02-NOV-1998; 98US-00184418.

PA (UABR-) UAB RES FOUND.

PI Hahn BH, Shaw GM, Gao F;

DR WPI; 2000-365651/31.

PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.

PS Claim 41; Fig 14; 131pp; English.

CC The present in invention provides the protein and coding sequences for  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of  
CC in a sample and to produce antibodies against non-subtype B HIV-1  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)

**SQ** Sequence 491 AA;

### Alignment Scores:

**Score: 2405.50**

Best Local Similarity: 91.26%

DB: 3

US-09-475-704A-3 (1-1479) X AAB6

QY 1 ATGGGCGCCGCCAGC

Db 1 MetGlyAlaArgAlaSer

61 CTGCCGCCCGCGGCAAG

Matches:	449
----------	-----

Mismatches: 16

Gaps: 1

1-491)

CGCGGCGGCAAGCTGGACG

ArgGlyGlyLysLeuAspA

QY	1	ATGAGGAGCCCGGCGCAGGATCCCTGGCGCGCGCGCAGGCTGGACGCGCTGGAGACGATCCG	60
		.....	
Db	1	MeGcYAlaAraAlaSerIleLeuAraGcYgLYuSLeuAraAraGTrpIduIuYrIeArg	20
QY	61	CTGCGCCCGCGCGCGCAGAAAGTGTACATGAATGAAGCACTGGTGTGGCCAGCGCGGAG	120
		.....	
Db	21	LeuAraProGcYgLYuSLeuYshIeYrMetIleYshIeLeuValTrpAlaSerAraGlu	40
QY	121	CTGAGAAAGTTGCGCTCTGAACCCCGCGCTGTGGAGACAGCGAGCGCTGCACCAATC	180
		.....	
Db	41	LeuGluAraPheAlaLeuAraAraProGcYleuLeuGluTrnSerAraGcYgLYuGlnIle	60

[illegible]

Db	420	AspCvrtHngIuArgInIaAenPheLeuGlyLysIleTrpProSerHsIuYsGlyArg	439
Qy	1331	CCGGCAACTTCCTGAGAGCCGCCCGAGGCCACCCGCCCGCCCGAGAGACTTCCGC	138
Db	440	ProGlyAenPheLeuGlnAenAArgProGluProThrAlAProProIaGluSerPheArg	459
Qy	1381	TTGAGAGAGACACACCCCGCCAGAGACGAGAGAGAGACGACCGCGAGACCTGACACG	144
Db	460	PheYsGluTrHrThProIaProYsGlnGluSerLysAaPargGluProLeuThrSer	479
Qy	1441	CTGAAGAGCGCTGTCGGCAACGACCCCTGAGCCAG	1476
Db	480	LeuYsSerIeuPheGlySerAspProLeuSerGln	491
RESULT 15			
ID	ADCT2913	standard; protein; 491 AA.	
XX	ADCT2913;		
AC	ADCT2913;		
DT	18-DEC-2003	(first entry)	
XX			
DE	HIV-1 gag protein containing an NCP sequence, SEQ ID NO 143.		
XX			
OS	complex; HIV nucleocapsid protein 7; NCP7; HIV-psi-site; anti-HIV;		
KW	human immunodeficiency virus; HIV.		
XX	Human immunodeficiency virus 1.		
PN	WO2003060098-A2.		
XX			
PD	24-JUL-2003.		
XX			
PF	10-JAN-2003; 2003WO-US000801.		
XX			
PR	11-JAN-2002; 2002US-0347369P.		
XX			
PA	(ACHI-) ACHILLION PHARM INC.		
PI	Beuchter D, Hou X, Marlor CW, Rice WG, Yang W;		
XX			
DR	WPI; 2003-646042/61.		
XX			
PT	Method of determining whether a compound inhibits formation of complex		
PT	between HIV nucleocapsid protein 7 polypeptide and HIV psi-site		
PT	oligonucleotide by comparing amount of complex formed in presence/absence		
PT	of compound.		
XX			
PS	Claim 2; SEQ ID NO 143; 105pp; English.		
XX			
CC	The invention relates to a novel method of determining whether a compound		
CC	inhibits formation of a complex between HIV nucleocapsid protein 7 (NCP7)		
CC	polypeptide and an HIV-psi-site oligonucleotide. The method involves		
CC	adding the HIV-psi-site oligo to a mixture of the NCP7 polypeptide and		
CC	the novel compound and comparing amount of complex formed, with that		
CC	formed in the absence of the compound, where a decrease in the amount of		
CC	complex formed in presence of the compound indicates that the compound		
CC	inhibits complex formation. The method of the invention involves anti-HIV		
CC	activity. The compounds identified are useful for treating a subject		
CC	infected with human immunodeficiency virus (HIV) by administering the		
CC	compound to the subject. This sequence represents an HIV-1 gag protein		
CC	which contains an NCP sequence of the invention.		
XX			
XX	Sequence 491 AA;		
Alignment Scores:			
Pred. No.:	1,246-134	Length:	491
Score:	2396.50	Matches:	449
Percent Similarity:	95.53%	Conservative:	21
Best Local Similarity:	91.26%	Mismatches:	21
Query Match:	7	Indels:	1
DB:		Gaps:	1

US-09-475-704A-3 (1-1479) x ADCT72913 (1-491)

QY	1	ATGGGCGCCGCGCCAGATCTCTGCGCGCGCGCAAGCTGACCGCTGGAGACGATCCGC	60
Db	1	MetGlyAlaArgAlaSerIleLeuArg961yAllylsLeuAspLysTrpGluLysIleArg	20
QY	61	CTGGCGCCCGCGCGCAAGAAGTGCATCATATGTAAGAAGCACTGGAGTGGGCGCAGCGCAG	120
Db	21	LeuArgProGlyIlyLysLysHisTyrMetIleLysHisLeuValTrpAlaSerArgGlu	40
QY	121	CTGGAGAAGTTCGCGCTTGAAACCCCGCGCTGCGAGAACACGAGCGCTGCAACGATC	180
Db	41	LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluTrpHisSerGluGlyCysLysGluIle	60
QY	181	ATCCGCGCAGCTGCACCCCGCGCTTGACACCGCGCAGCGAGAGCTGAGAGCGCTTCAAC	240
Db	61	IleLysGlnLeuGlnProAlaLeuGlnThrGlyLysGluLeuAspSerLeuHisAsn	80
QY	241	ACCGTGGCCACCTGTACTGTGCGTACAGAGAAGATCGAGTCCGCGCAGACCAAGAGAGCC	300
Db	81	ThrValAlaTrpLeuTyrCysValHisAlaGluIleGluIleArgAspThrLysGluAla	100
QY	301	CTGGACAAGATCAGAGAGAGACAGACAAGATGCCAGACAGAAATCCAGACCGCAGGCC	360
Db	101	LeuAspLysIleGluGluGluGlnGluLysSerGlnGluLysTrpGlnGlnAlaLysGlu	120
QY	361	GCCGACAAGGCGCAAGTGAAGCCACAATACCCCATCGTGCAGAACTGTACAGGCCAGATG	420
Db	121	AlaAsp--GlyLysValSerGlnAsnTyrProIleValGlnAsnLeuGlnGlyMet	139
QY	421	GTGACACCAAGGCATCAGCGCCCGCGCACCTGAAGCGCTGGAGAGGTATCGAGAGAAG	480
Db	140	ValHisGlnAlaLeuSerProArgThrLeuAsnAlaTrpValLysValIleGluLys	159
QY	481	GCCTTCAGCGCCGAGGTGATCCCATGTTCAACGCGCTGAGCGAGCGCGCACCCCGCAG	540
Db	160	AlaPheSerProGluIleIleProMetPheThrAlaLeuSerGluGlyAlaTrpProGln	179
QY	541	GACCTGAACACGATGTTGAACAACCTGTGGCGCGCCACACGAGCGCGCATTCAGATCTGAG	600
Db	180	AspLeuAsnThrMetLeuAsnThrValGlyHisGlnAlaAlaMetGlnMetLeuLys	199
QY	601	GACACCATCAACAGAGAGAGCGCGCAGTGGGAGACCGCGTCAACCGCTGCACGCGCGCCC	660
Db	200	AspThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisProAlaGlnIleGlyPro	219
QY	661	ATGCGCCCGCGCAGATGCGGAGCGCCCGCGGACGACATGCGCGGACCAACCAAGCAC	720
Db	220	IleAlaProGlyGlnMetArgGluProArgLysSerAspIleAlaGlyThrThrSerThr	239
QY	721	CTGCAGAGAGCATCGCTGTGATATCACAGCAACCCCGCATCCCGCTGGGAGCATCTAC	780
Db	240	LeuGlnGluGlnIleAlaTrpMetThrGlyAsnProProValGlyGluIleLys	259
QY	781	AAGCGGTGATCATCTGGGCGCTTAAACAAGATCGTGGATGTACAGCGCGTAGCATC	840
Db	260	LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetCysSerProValSerIle	279
QY	841	CTGCACATCAAGAGGCGCCCAAGAGCGCTTCGCGCATCGTGAACCGCTTCTCAAG	900
Db	280	LeuAspIleLysGlnGlyProLysGluProPheArgAspTyrValaAspArgPheLys	299
QY	901	ACCTGGCGCGCGAGCAGAGACCCACAGAGGTAAAGATCGATGACCGACCACTGTG	960
Db	300	ThrLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetTrpAspThrLeuLeu	319
QY	961	GTGACAAAGCGCAACCGCATGCAACAACATCTGCGCGCTTCGCGCGCGCGCGCAGC	1020
Db	320	ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer	339
QY	1021	CTGAGAGAGATGATGACCGCGCTGCACAGCGCGTGGCGCGCGCCACGACCAAGCGCGCGTG	1080

```
Db      340 LeuGIuMeMeTThrAlaCysGlnGlyValGlyGlyProSerHisIleValArgVal 359
QY      1081 CTGGCCGAGGCGATAGCCAGGCCAAACACAGCGTGAATGACAGAAAGCAACTTCAAG 1140
Db      360 LeuAlaGluAlaMeSerGlnThrAsnAsnSerIleLeuMeGlnArgSerAsnPheLys 379
QY      1141 GGGCCCCGGCGATGATGTCAGTGCCTTCAACTGGGCAAGAGGGCCACATGGCCGGCAAC 1200
Db      380 GlyPheLysArgThrValIleCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsn 399
QY      1201 TGCCCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGGCAAGAGGGCCACAGATGAAG 1260
Db      400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 419
QY      1261 GACTGCACCGAGCGCCAGGCCCACTTCTGGGCAAGATCTGGCCCAAGCCCAAGGGCGGC 1320
Db      420 AspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439
QY      1321 CCCGGCACTTCTGCAAGCGCGCCCGAGCCCAAGCCCGCCCGCGCGAGCTTCCGC 1380
Db      440 ProGlyAsnPheLeuGlnAsnArgProGluProThrAlaProProIleGluSerPheArg 459
QY      1381 TTCGAGAGACCAACCCCGGCCAGAGACAGAGAGCAAGACCGCAGACCTTGACCAAGC 1440
Db      460 PheGluGluIuThrThrProAlaLeuLysGlnGluIuLysAspArgGluProLeuThrSer 479
QY      1441 CTGAAGAGCGCTTGTGGGCAAGCAACCCCTGAGCCAG 1476
Db      480 LeuLysSerLeuPheGlySerAspProLeuSerGln 491
```

Search completed: March 11, 2005, 15:22:11  
Job time : 242.694 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 15:04:55 ; Search time 51.7254 Seconds  
(without alignments)  
5502.310 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 2772  
Sequence: 1 atggcgccgcgcgcagcatc.....acgaccctcagcagcagtaa 1479

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO\_spool\_p/US09475704/runatc 10032005 140222 14839/app\_query.fasta\_1.3342  
-DB=PIR 79 -QFMT=faastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0 -LIST=45  
-UNITS=BITS -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdt -LST=45  
-DOCALLIGN=200 -THR.SCORE=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=PCO -NORM=EXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09475704\_QCGN 1.1.160 @runatc 10032005 140222 14839 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOBVERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGIO  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
PIR 79:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	2300.5	83.0	501	2	SS4377 gag polypotein -
2	2294.5	82.8	497	1	FOJLND gag polypotein -
3	2266.5	81.8	505	2	TO1667 gag polypotein -
4	2263.5	81.6	500	1	FOVWLV gag polypotein -
5	2249	81.1	500	1	FOVWHA gag polypotein -
6	2247	81.1	502	1	FOVWA2 gag polypotein -
7	2235	80.6	512	1	FOVWHA3 gag polypotein -
8	2233	80.6	498	2	TO9436 gag polypotein -
9	2231	80.5	500	1	A44001 gag polypotein -
10	2228	80.4	500	2	S33979 gag polypotein -
11	2216.5	80.0	506	1	A38068 gag polypotein -
12	2037.5	73.5	478	1	FOVWVL gag polypotein -
13	1989	71.8	508	1	FOJLST gag polypotein -
14	1400	50.5	521	2	S08435 gag polypotein -

15	1399.5	50.5	521	2	SS3091 gag polypotein -
16	1398	50.4	521	1	FOJLST gag polypotein -
17	1392.5	50.2	554	1	S46346 gag polypotein -
18	1380	49.8	521	1	FOJLTA gag polypotein -
19	1376	49.6	522	1	FOJLGG gag polypotein -
20	1375	49.6	521	2	S12152 gag polypotein -
21	1371	49.5	507	2	T11559 gag polypotein -
22	1370	49.4	507	2	S04237 gag polypotein -
23	1364.5	49.2	522	1	FOJLJ2 gag polypotein -
24	1364	49.2	510	1	FOJLTM gag polypotein -
25	1354.5	48.9	506	1	FOJLJ3 gag polypotein -
26	1354	48.8	519	1	FOJLJ4 gag polypotein -
27	1341.5	48.4	506	1	FOJLJ5 gag polypotein -
28	1222.5	44.1	502	2	S28080 gag polypotein -
29	716.5	25.8	171	2	SS2929 gag polypotein -
30	692	25.0	146	2	S60708 gag polypotein -
31	688	24.8	146	2	S60698 gag polypotein -
32	688	24.8	146	2	S60704 gag polypotein -
33	688	24.8	146	2	S60697 gag polypotein -
34	683	24.6	146	2	S60699 gag polypotein -
35	683	24.6	146	2	S60702 gag polypotein -
36	680	24.5	146	2	S60700 gag polypotein -
37	680	24.5	146	2	S60703 gag polypotein -
38	560	20.2	212	2	S03070 gag polypotein -
39	454.5	16.4	450	1	FOJLJP gag polypotein -
40	453.5	16.4	449	2	S23819 gag polypotein -
41	449	16.2	449	2	A45557 gag polypotein -
42	445	16.1	486	1	FOJLJEV gag polypotein -
43	441	15.9	85	2	S49086 gag polypotein -
44	435.5	15.7	450	2	S25162 gag polypotein -
45	431.5	15.6	1106	2	J00405 hypothetical 119.5

## ALIGNMENTS

## RESULT 1

SS4377  
gag polypotein - human immunodeficiency virus type 1  
C/Species: human immunodeficiency virus type 1, HIV-1  
C/Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004  
R/Theodore, T.; Buckler-White, A.J.  
C/Accession: SS4377  
submitted to the EMBL Data Library, July 1989  
A/Reference number: SS4377  
A/Accession: SS4377  
A/Status: preliminary  
A/Molecule type: genomic RNA  
A/Residues: 1-501 <THR>  
A/Cross-references: UNIPROT:P12495; EMBL:M2639; NID:g329377; PIDN:AAA4565.1; PID:g3293;  
C/Superfamily: AIDS-related virus gag polypotein  
C/Keywords: polypotein

## Alignment Scores:

Pred. No.:	1.03e-109	Length:	501
Score:	2300.50	Matches:	434
Percent Similarity:	93.21%	Conservative:	33
Best Local Similarity:	86.63%	Mismatches:	25
Query Match:	82.99%	Indels:	9
DB:	2	Gaps:	4

US-09-475-704A-3 (1-1479) x SS4377 (1-501)

QY	1	ATGGCGCCCGCGCCAGCATCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCG	60
DB	1	MetGlyAlaArgAlaSerValLeuSerGlyLysLeuAlaAlaTrpGluLysIleArg	20
QY	61	CTGGCGCCCGCGCGGAGAGAGTCATGATGAGAGCACTGTGTGGGCGCGCGAG	120
DB	21	LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys	40
QY	121	CTGGAAGATTGCGCTGAGACCCCGCGCTGTGTGAGAGACGAGCGGCTGCAAGATC	180
DB	41	LeuGluArgPheAlaLeuAsnProGlyLeuGluLeuGluLutTrpSerAspGlyLysGlnIle	60

```
QY 181 ATCCGAGCTGACCCCGCTGACAGCCGCGAGGAGGAGCTTGAAGCTTGTTCAC 240
Db 61 TleGlyInleuInProAlaIleArgThnGlySerGluInleuSerLeuPhean 80
QY 241 ACCGTGGCCACCTGTACTGCTGACAGAGAATGAGGTCCGACACCAAGAGGCC 300
Db 81 ThrValAlaThrleuYrCyValHsGluArgIleGluValIlyAspThrIysGluAla 100
QY 301 CTGGACAAGATCGAGAGAGAGCAACAAGTCCACAGAAAGATCCAGAGCCGAGGCC 360
Db 101 LeuGluIlyMetGluGluGluInleuInleuSerIlyAsnIlySerIlyGluInleuAla 120
QY 361 GCCGACAGAGGC-----AGGTGAGCCACAATACCCCTCTGTGAGAACCTTGAC 411
Db 121 AlaAspAlaGlyAsnAsnSerGluValSerGlnAsnIlyProIleValGlnInleuGln 140
QY 412 GGCAGATGATGACAGGACCATCAGCCCGCAGCCCTGACAGCTGGTGAAGTATC 471
Db 141 GlyGlnMetValHsGlnAlaIleSerProArgThrLeuAsnAlaTrpValIlyValIle 160
QY 472 GAGAGAGAGCTTCAAGCCCGAGGTGATCCCATTTCAAGCCCTGAGAGGAGGCC 531
Db 161 GluGluIlyAlaIlePheSerProGluValIleProMetPheSerAlaIleuSerGluGlyAla 180
QY 532 ACCCCCGAGGACCTGAACAAGATGTTGAACACCTGCGGAGGAGCCAGCATGAC 591
Db 181 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHsGlnAlaIleMetGln 200
QY 592 ATGCTGAAGAGACACCATCAACAGAGAGGCGCGAGTGGAGACCGCTGACCCCTGCAC 651
Db 201 MetLeuIlyGluThrIleAsnGluGluIleAlaIleGluTrpAspArgLeuHsIleProValHs 220
QY 652 GCGGCGCCATGCGCCCGGCGAGAGCGCCGAGCCCGGCGAGGACATCCCGGACCC 711
Db 221 AlaGlyProIleAlaIleProGlyGlnMetAspGluProArgIlySerAspIleAlaGlyThr 240
QY 712 ACCAGACCCCTGACAGAGAGATGCGCTGATGACAGACACCCCGGCGGAGGCC 771
Db 241 ThrSerThrleuGlnGluInleuAlaIleTrpMetThrSerAsnProIleProValGly 260
QY 772 GACATTCACAGCGGTGATCATCTCGGCGCTGAACAAGATGTCGGATGTACAGCCC 831
Db 261 GluIleYrIlyArgTrpIleIleleuGlyLeuAsnIlyIleValArgMetIlySerPro 280
QY 832 GTGAGATCTGTGACATCAAGAGGCGCCGAGAGGACCTTCCGCGACATCTGACGCC 891
Db 281 ValSerIleLeuAspIleArgGlnGlyProIlyGluProPheArgAspIlyValAspArg 300
QY 892 TTCTTCAAGACCTGCGCGCGAGAGAGACACCCAGAGAGTGAAGAACTGACCGAC 951
Db 301 PheYrIlySerThrleuArgAlaGluGlnAlaSerGlnGluValIlyGlyYrIlePheThrGln 320
QY 952 ACCGTGCTGTGACAGACCGCAACCCCGACCTGCAAGAACCATCTGCGGCTTCGCCCC 1011
Db 321 ThrleuLeuValGlnAsnAlaAsnProAspCyIlySerIleleuIlyAlaIleuGlyPro 340
QY 1012 GCGCGACGCTGAGAGATGATGACCGCTGCGGAGGCGGCGGCGCCGACCAAG 1071
Db 341 GlnAlaThrleuGluIleuMetThrIlyAcysGlnGlyValGlyIlyProSerIlyAsn 360
QY 1072 GCCCGCTGTGCGCGAGGCGATGACAGGCGCAACACAGC-----GTGATGATG 1122
Db 361 AlaArgValIleuAlaGluAlaMetSerGlnAlaThrAsnSerAlaAlaAlaValMetIle 380
QY 1123 CAGAAAGCAACTTCAAGAGGCGCCCGCGCATCTCAAGTGTCACTGCGGCAAGAG 1182
Db 381 GlnArgGlyAsnPheIlyGlyProArgIlySerThrIleIlyCyPheAsnCyGlyIlyGlu 400
QY 1183 GGCACATCGCGCGCACTGCGGCGCGCGCGCGAGAGAGGCGTGTGAAATGCGGCGAG 1242
Db 401 GlyHsIleAlaIlyAsnCyAspArgAlaProArgArgIlyGlyCySerIlyIlySerIlyIly 420
```

```
QY 1243 GAGGCGCACAGATGAAGATCTGACCCGAGCGCCAGCCCACTTCTGGGCAAGATCGG 1302
Db 421 GluGlyHsGlnleuIlyAsnIlyAspCySerThrGluArgGlnAlaAsnPheIleuGlyIlySerIle 440
QY 1303 CCCAGCGCAAGAGGCGCGCGCGCGCAACTTCTCGAGAGAGCGCGCGCGCGCGCGCC 1362
Db 441 ProSerHsIlyGlyIlyArgProGlyAsnPheIleuGlnSerArgProGluProThrAlaPro 460
QY 1363 CCCGCGAGAGCTTCCGCTTC---GAGAGACACACCCCGCGCAGAGCAGAGAGCAAG 1419
Db 461 ProAlaGluSerPheGlyIlyPheGlyGluIleThrProSerGlnIlyGlnIlyIly 480
QY 1420 GACCGCGAGACCCCTG-----ACGACCTGAAAGACCTTGTGGCAAGACCCCTGAC 1473
Db 481 AspIlyGluIleuYrProSerThrAlaIleuIlySerIlePheGlyAsnAspProIleu 500
QY 1474 CAG 1476
Db 501 Gln 501
```

## RESULT 2

```
FOIJD
gag polyprotein - human immunodeficiency virus type 1 (isolate NDK)
N.Alternate names: core polyprotein
N.Contants: core protein p15; core protein p17; core protein p24
C.Species: human immunodeficiency virus type 1, HIV-1
A.Note: host Homo sapiens (man)
C.Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C.Accession: J00065
R.Splice, B.: Site, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.,
Gene 81, 275-284, 1989
A.Title: Nucleotide sequence of HIV-1-NDK: a highly cytopathic strain of the human immuno-
A.Reference number: J00065; MUID:90034200; PMID:2806917
A.Accession: J00065
A.Molecule type: DNA
A.Residues: 1-497 <SPI>
A.Cross-references: UNIPROT:P18800; GB:M27323; NID:G328154; PIDN:AA44868.1; PID:G328157
C.Genetics:
A.Gene: gag
C.Superfamily: AIDS-related virus gag polyprotein
C.Keywords: AIDS; core protein; immunodeficiency; polyprotein
F.1-129/Product: core protein p17 #status predicted <C17>
F.130-389/Product: core protein p24 #status predicted <C24>
F.390-497/Product: core protein p15 #status predicted <C15>
```

## Alignment Scores:

Pred. No.:	2.08e-109	Length:	497
Score:	2294.50	Matches:	428
Percent Similarity:	93.57%	Conservative:	38
Best Local Similarity:	85.94%	Mismatches:	25
Query Match:	82.77%	Indels:	7
Db:	1	Gaps:	4

US-09-475-704A-3 (1-1479) x FOIJD (1-497)

```
QY 1 ATGGCGCGCGCGCGAGATCTGCGCGCGCAAGCTGACGCGTGGAGCGATCCGC 60
Db 1 MetGlyAlaArgAlaSerValIleuSerGlyIlyIlyAsnIlyThrTPGluArgIleArg 20
QY 61 CTGGCGCGCGCGCGCAAGATGTCTACATGATGAAGACCTGTGTGGCGCGCGAG 120
Db 21 LeuArgProGlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIly 40
QY 121 CTGAGAAATTCGCGCTGAACCCCGCGCTGTGAGACCAAGCGGCGCTGCAACAGATC 180
Db 41 LeuGluArgPheThrleuAsnProGlyIleuIleuIlyThrSerGluIlyCyIlyGlnIle 60
QY 181 ATCCGACAGTCGACCCCGCGCTGACAGCGGCGAGAGAGCTGAAGAGCTTCAAC 240
Db 61 TleGlyInleuGlnProSerIleGlnThrIlySerIleGluIlyIlyIlyIlyIlyIly 80
QY 241 ACCGTGCGACCCCTGTACTGCTGACAGAGAGTCAAGATTCAGGTCGCGAGAGGCC 300
```

D	b	81	ThrValAlaIleThrLeuTyrCysValHisGluArgIleGluValValAspThrLysGluAla	100
Q	y	301	CTGGACACAGATGAGGAGGAGCAAGAACAGTCCGACGACAGAGATCCAGAGCCGAGCC	360
D	b	101	ValGlnTyrMetGlnGluGlnGlnAsnLysSerTyrLysIleThrGlnGlnAla--Ala	119
Q	y	361	GCCGCAAGAGGGCAGGTGAGCCAGAACCTAACCTGTCAGAACCTGCAAGGGCCAGATG	420
D	b	120	AlaAspSerSerGlnAlaSerGlnAsnTyrProIleValGlnAsnLeuGlnGlnMet	139
Q	y	421	GTGCACCAAGGCCATCAGCCCGCAGCCCTTGAAAGCCCTGGGGTGAAGGTGATTCGAGAGAAG	480
D	b	140	ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIleGluGlnLys	159
Q	y	481	GCTTTCAGCCCGCAGGTGATCCCATGTTTCACCGCCCTGAGCGAGGGCCACCCCCAG	540
D	b	160	AlaPheSerProGluValIleProMetPheSerAlaLeuSerGlnGluYalaTrpProGln	179
Q	y	541	GACCTGAACACAGATGTTGAACACCGTGGGGCGGCACAGAGCCGCATGAGATGCTGAAG	600
D	b	180	AspLeuAsnThrMetCysLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys	199
Q	y	601	GACCAACATCAACGAGAGAGGCCCGCCAGTGGAGACCGCGTACACCCCGTGCACGCCGCC	660
D	b	200	GlnThrTrpIleAsnAspGlnAlaIleGluTrpAspArgLeuHisIleProValHisIleArgLysPro	219
Q	y	661	ATCGCCCGCCGCGCAGATGCGCAGAGCCCGGCGAGCGACATGCGCGGCACCAACAGACACC	720
D	b	220	ValAlaIleProGlnGlnMetArgGlnProArgLysSerAspIleAlaGluYthrThrSerThr	239
Q	y	721	CTGCAGAGAGCAGATGGCGCTTGATGATACACAGAACCCCCCATCCCGTGGCGACATCTAC	780
D	b	240	LeuGlnGlnGlnGlnIleAlaTrpMetThrSerAsnProProIleProValGlyGluIleTyr	259
Q	y	781	AAGCGGTGATCATCTCTGGGCTTGAAACAAGATCGTCCGATGACAGCCCGGAGACATC	840
D	b	260	LysArgTrpIleIleLeuGlnYleuAsnLysIleValArgMetLysSerProValSerIle	279
Q	y	841	CTGCACATCAAGCAGAGGCCCGCCAGAGACCCTTCCGCGACTACGTGAGACCGCTTCTTCAAG	900
D	b	280	LeuAspIleArgGlnGlnYProLysGlnProPheArgAspTyrValAspArgPheTyrLys	299
Q	y	901	ACCTTCGCGCGCGAGCAGACGACCCAGAGAGTGAAGACCTGATGACCGACACCTGTCTG	960
D	b	300	ThrLeuArgAlaGlnGlnAlaSerGlnAspValLysAsnTrpMetThrGlnThrLeuLeu	319
Q	y	961	GTGCAGAACCGCCACCGCCGCTGCAAGAACATCTCGCGGCTCTCGCGCCCGCGCGAC	1020
D	b	320	ValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGluYProGlnAlaThr	339
Q	y	1021	CTGCAGAGAGATGATGACCGCTGCGCAGAGCGCGTGGCGCGCCAGCCACAGAGCCCGCGTG	1080
D	b	340	LeuGlnGlnMetMetThrLysCysGlnGlnYValGlyGlyProGlnYHisLysValAlaArgVal	359
Q	y	1081	CTGCGCGAGCGATGAGCCAGGCCAAC-----ACCAGCGTGAATGACAGAAAGCC	1133
D	b	360	LeuAlaGlnAlaMetSerGlnValThrGlySerAlaThrAlaValMetMetGlnArgGly	379
Q	y	1132	AACCTCAAGAGGGCCCCCGCGCATCGCAAGTCTTCAATGCGGCAAGAGAGGGCCACATC	1199
D	b	380	AsnPheLysGlyProArgLysSerIleLysCysPheAsnCysGluYlySerGlnGlnThr	399
Q	y	1192	GCCCGCAACTGCGCGCGCCCCCGCCAGAGAGGGCGCTGTGAGATGTCGCGCAAGAGAGGCCAC	1255
D	b	400	AlaLysAsnCysArgAlaIleProArgLysLysGlyCysTrpLysCysGluYArgGlnGlnHis	419
Q	y	1252	CAGATGAAGAGACTGCACCGAGCGCCAGGCCAACCTTCTGGGCAAGATCTGCGCCACGCCAC	1311
D	b	420	GlnMetLysAspCysThrGlnArgGlnAlaAsnPheLeuGlnLysIleTrpProSerHis	439
Q	y	1312	AAGGGCGCGCCCGCGGACCTTCTTGCAGAGCCCGCCAGGCCACCGCCCCCGCGGAG	1377
D	b	440	LysGlnYArgProGlnYAsnPheLeuGlnSerArgProGlnProThrAlaProAlaGln	459

**RESULT 3**

gag polyprotein - human immunodeficiency virus type 1  
C.Species: human immunodeficiency virus type 1, HIV-1  
C.Date: 19-Feb-1999 #sequence\_revstion 19-Feb-1999 #text\_change 09-Jul-2004  
A.Accession: T01667  
R.Allison, M.; Main-Hobson, S.; Gluckman, J.C.; Sonigo, P.  
Call 46, 63-74, 1986  
A.Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol  
A.Reference number: Z14389; MWID:86245056; PMID:2424612  
A.Accession: T01667  
A.Status: preliminary; translated from GB/EMBL/DDBJ  
A.Molecule type: mRNA  
A.Residues: 1-505 <ALT>  
A.Cross-references: UNIPROT:P04594; EMBL:K03456; NID:g60228; PIDN:CAA28011.1; PID:g60222D  
C.Subfamily: AIDS-related virus gag polyprotein

**Alignment Scores:**

Pred. No.:	5	51e-108	Length:	505
Score:	2266.50	Matches:	431	
Percent Similarity:	90.71%	Conservative:	28	
Best Local Similarity:	85.18%	Mismatches:	32	
Query Match:	81.76%	Indels:	15	
	2	Gaps:	5	

US-09-475-704A-3 (1-1479) x T01667 (1-505)

DY 1 ATGGGCGCCGCCGCACATCTCGCGCGCGCAAGCTGAACGCTTGAGAGCATCCGC 60

Db 1 MetGIyAlaArgAlaSerValIleuSerGIyGIyLeuAepAlaTrpGIuIuysrIlearg 20

DY 61 CTGGCGCCCCGGGGGCAAGAAGTCTACATGATGAAGCACTGTGTGGGCCAGCCGGAG 120

Db 21 LeuaAgPrOGlIGlyLyAsylsYsrYrArgLeuIushIsuevaITrpAlaserArglu 40

DY 121 CTGGAAGAATGGCCCTGAACCCCGGCTGCTGAGACCAAGCGGGGCTGCAAGCAGATC 180

Db 41 LeuGIuArpPheAlaIeuasnPrOGlyLeuIeuGIuThrgIyGIuGIyCysglndnIle 60

DY 181 ATCCGCCAGCTCACCCCGCCCTGACAGACCGGACGCGAGCAGAGACTGAAGAGCCTGTTCAC 240

Db 61 MetGIuGIuInIeuGIuSerThrIeuYsrThrgIySergIuGIuIuIeYserIeuYsrhen 80

DY 241 ACCGTGGCCACCTCTGACTGCGCTGCGACAGAAAGTCAAGTCCGCGACACCAAGAGGCC 300

Db 81 ThrValAlaThrLeuYrCySvalIHsglnArgIleaSpValIysAspThrIysgluIa 100

DY 301 CTGGAACAAGATGAGAGGAGGAGCAACAAGTGCAGAGAGAATCCAGCAGCGCAGGCC 360

Db 101 LeuaepIysrIleGIuGIuIuIleGIuInIeuIysSerArgIuGIuYsrThrgIuInIalala 120

DY 361 GCC-----GACMAAGGCAAGGTGAGCCAGACTAACCCCATC 396

Db 121 AlaGIuInIalAlaIalAlaIaThrIysAsnSerSerValISerGIuAnsrYrProile 140

DY 397 GTGCAAGAACTGCAAGGCGCAGATGTGCAACCAGGCCATCAAGCCCCCGACCTTGAAACC 456

Db 141 ValGIuAnsrAlaaglndIyglInMetIleHsglnAlaIleSerProAgyThrLeuasnla 160

DY 457 TGGGGAAGATGATGAGAGAGAAAGGCTTCAGGCCCGCAGAGTGATCCCATTGCAACGCC 516

Db 161 TrpValIysValIleGIuGIuIuYsalIapheserPrOGluValIIlePrometPheserAla 180

DY 517 CTGACCGAGGGCGCCACCCCGCAGACTGAACAGATGTTGAACACCGTGGCGCGCAC 576

```
Db      ||| 181 LeuSeGluGlyAlaThrProGlnAspLeuAsnMetLeuAsnIleValGlyHis 200
Qy      ||| 577 CAGGCGCCATGCAAGTCTTGAAGACACCATCAACGAGAGCGCGGAGTGGAGCCG 636
Db      ||| 201 GlnAlaIleMetGlnMetLeuAspThrIleAsnGlnGlnAlaAlaAspTrpAspArg 220
Qy      ||| 637 GTGCAACCCCGTGCAGCGCGCCCATCGCCCGGCAATGCGGAGCGCGCGGAGC 696
Db      ||| 221 ValHisProValHisAlaGlyProIleProGlnProGlnMetArgGlnProArgGlySer 240
Qy      ||| 697 GACATGCGCGGACCAACGAGCCCTGAGAGAGAGATGCGCTGATGACCAACCC 756
Db      ||| 241 AspIleAlaGlyThrThrSerThrLeuGlnGlnIleGlyTrpMetThrSerAsnPro 260
Qy      ||| 757 CCCATCCCGTGGGAGCATCTTCAACGCGGTGATCATCTGGCGCTGAACAAGTCTG 816
Db      ||| 261 ProIleProValGlyAspIleTrpAspArgTrpIleIleLeuGlyLeuAsnIleVal 280
Qy      ||| 817 CGGATGTCAGCGCGGTGAGCATCCCTGACATCAACGAGGCGCGCAAGGCCCTTCG 876
Db      ||| 281 ArgMetTrpSerProValSerIleLeuAspIleArgGlnGlyProIleGlnProAspArg 300
Qy      ||| 877 GACTAGTGAACCGCTTCTTCAAGACCTGCGCGCGGACGAGACCAACGAGGTGAAG 936
Db      ||| 301 AspTrpValAspArgPhePheThrLeuArgAlaGlnAlaThrGlnIleValHis 320
Qy      ||| 937 AACTGATGATCCGACACCTCTGTGTGCAACGCCCAACCCGACTGCAAGCAATCTTG 996
Db      ||| 321 AsnTrpMetThrGlnThrLeuLeuValGlnAsnAlaAspProAspCysLeuThrIleLeu 340
Qy      ||| 997 GCGGCTCTGCGCGCGCGGACGCTGAGAGAGATGATGACCGGCTGCGCGCGTGGCG 1056
Db      ||| 341 LysAlaLeuGlyProGlyAlaThrLeuGlnGlnMetMetThrAlaCysGlnIleValGly 360
Qy      ||| 1057 GCGCCGACGACAGGCGCGCGTGTGTGCGGAGGAGATGAGCGACGACCAACCGAC 1113
Db      ||| 361 GlyProSerHisLeuAlaArgValLeuAlaGlnIleAsnSerGlnAlaThrAsnSerThr 380
Qy      ||| 1114 -----GTGATGATGCAAGAGCACTTCAAGGCGCGCGCGCATGCTGATGCTTC 1167
Db      ||| 381 AlaAlaIleMetMetGlnArgGlyAsnPheLeuGlnIleValArgIle----LysCysPhe 399
Qy      ||| 1168 AACTGCGGACGAGAGGCGCACATCGCGCGGACATGCGCGCGCGGACGAGAGGCGTGC 1227
Db      ||| 400 AsnCysGlyLysGlnGlyHisLeuAlaAsnCysAspAlaArgArgLysLysCys 419
Qy      ||| 1228 TGGAAATGCGGACGAGGCGCACAGATGAAGACTGACCGAGCGCGGACCAATCTTC 1287
Db      ||| 420 TrpLysCysGlyLysGlnGlyHisLeuMetLysAspCysThrGlnArgGlnAlaAsnPhe 439
Qy      ||| 1288 CTGGGCAAGATGTGGCGCCAGCCCAAGAGGCGCGCGCGGACATCTTCTGAGAGCGCGCC 1347
Db      ||| 440 LeuGlyLysIleTrpProSerHisLysGlyArgProGlnAsnPheLeuGlnSerTrpPro 459
Qy      ||| 1348 GAGCCACCGCGCGCGCGCGGAGCTTCGCGCTTC---GAGGAGACCAACCGCGCGGACG 1404
Db      ||| 460 GlnProThrAlaArgProAlaIleAsnSerPheGlyPheGlyGlnGlnIleLysProSerGln 479
Qy      ||| 1405 AAGCAGAGAGAGCAGAGACCGGAG-----ACCTGACCAAGCTGAAGAGCTGTTCCGC 1458
Db      ||| 480 LysGlnGlnGlnIleLysAspLysGlnLeuTrpProLeuAlaSerLeuLysSerLeuPheGly 499
Qy      ||| 1459 AAGCAGACCCCTGAGCCAG 1476
Db      ||| 500 AsnAspGlnLeuSerGln 505
```

## RESULT 4

FOVWLTV  
gag polyprotein - human immunodeficiency virus type 1 (isolate LAV-1a)  
NAlternate names: assemblin; core polyprotein; gag precursor  
NContains: capsid antigen core protein p24CA; core protein p1; core protein p6; matrix  
CSpecies: human immunodeficiency virus type 1, HIV-1

```
A>Note: host Homo sapiens (man)
C>Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 05-Dec-1998
C/Accession: A03946
R/Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A/Title: Nucleotide sequence of the AIDS virus, LAV.
A/Reference number: A90866; MUID:85099333; PMID:2981635
A/Accession: A03946
A/Molecule type: DNA
A/Residues: 1-500 <MAI>
C/Genetics:
A/Gene: gag
C/Superfamily: AIDS-related virus gag polyprotein
C/Keywords: AIDS; blocked amino end; core protein; immunodeficiency; lipoprotein; myrist
F/2-500/Product: gag precursor (assemblin) #status predicted <GAG>
F/2-132/Product: matrix antigen core protein p17MA #status predicted <P17>
F/20-32/Region: nuclear location signal
F/110-114/Region: nuclear location signal
F/133-363/Product: capsid antigen core protein p24CA #status predicted <P24>
F/364-377/Product: core protein p2 #status predicted <CP2>
F/378-432/Product: nucleocapsid core protein p7NC #status predicted <CP7>
F/392-405/Region: zinc finger CCHC motif
F/413-426/Region: zinc finger CCHC motif
F/433-448/Product: core protein p1 #status predicted <CP1>
F/449-500/Product: core protein p6 #status predicted <CP6>
F/2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted
F/392-395,400,405/Binding site: zinc (Cys, Cys, His, Cys) #status experimental
F/413,416,421,426/Binding site: zinc (Cys, Cys, His, Cys) #status experimental

Alignment Scores:
Pred. No.: 8,29e-108 Length: 500
Score: 2263.00 Matches: 425
Percent Similarity: 92.61% Conservative: 39
Best Local Similarity: 84.83% Mismatches: 27
Query Match: 81.64% Indels: 10
DB: 1 Gaps: 5

US-09-475-704A-3 (1-1479) x FOVWLTV (1-500)
Qy      ||| 1 ATGGGCGCGCGCGCGCATCTTGGCGGCGGCAAGCTGAGCGCTGGAGCGCATCCG 60
Db      ||| 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyIleLeuAspArgTrpGlnIleArg 20
Qy      ||| 61 CTGGGCGCGCGCGCGCAAGTGTACATATGAAGCACTGGTGGCGCGGCGGAG 120
Db      ||| 21 LeuArgProGlyGlyLysLysLysTrpLysLeuLysHisIleValTrpAlaSerArgGln 40
Qy      ||| 121 CTGGAAGATTCGCGCTGMAACCGCGCGCTGTGAGACAGCGAGGCGCTGCAACAGATC 180
Db      ||| 41 LeuGlnArgPheAlaValAsnProGlyLeuLeuGlnThrSerGlnGlyCysArgGlnIle 60
Qy      ||| 181 ATCCGCAAGCTGACACCGCGCGCTGCAAGACCGGACGAGAGAGCTGAAGAGCTTGTAC 240
Db      ||| 61 LeuGlnLeuGlnProSerLeuGlnThrGlyGergGlnLeuAspSerLeuTrpAsn 80
Qy      ||| 241 ACCGTGGCAACCTGTCTGTGTCAGAGAAATGCAAGTCCGCGGACCAACCAAGAGGCC 300
Db      ||| 81 ThrValAlaThrLeuTrpCysValHisGlnArgIleGlnIleLysAspTrpLysGlnAla 100
Qy      ||| 301 CTGGAACAAGATCGAGAGAGCAGAACAGATGCCAGAGAGATTCAGAGGCGCGAGGCC 360
Db      ||| 101 LeuAspLysIleGlnGlnGlnIleAsnLysSerLysLysValGlnGlnAla---Ala 119
Qy      ||| 361 GCCGACAGGSC-----AAGTGAACCAAGACTACCCCATCTGTCAGAACCTGAG 411
Db      ||| 120 AlaAspThrGlyHisSerSerGlnValSerGlnAsnTrpProIleValGlnAsnIleGln 139
Qy      ||| 412 GGCAGATGTGTCACAGGCGCATGACCGCGCGCGGACCGCTGAACGCGTGGTGAAGTATC 471
Db      ||| 140 GlyIleMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValVal 159
Qy      ||| 472 GAGGAGAAAGGCTTCAGCGCGGAGGTATCCCATGTTTCAACGCGCTGAGCGAGGCGCC 531
```

Db 160 GluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGlyAla 179  
QY 532 ACCCCCGAGACCTTGAACAGATGTTGAACACCTGTGGCGGCCACCGCGCCATGCGAG 591  
Db 180 ThrProGlnPheLeuSerThrMetLeuSerThrValGlyIleGlnAlaAlaMetGln 199  
QY 592 ATGCTGAAGACACCATCAACAGAGAGCGCGAGTGGAGCCGCTGCACCCCGTGAC 651  
Db 200 MetLeuLysGluThrIleAsnGluGluAlaIleGluThrAspArgIleHisProValHis 219  
QY 652 GCCGCGCCCATGCGCCCGCGCCAGATGCGCGAGCCCGCGCGACGACATGCGCGCAC 711  
Db 220 AlaGlyProIleAlaProGlyGlnMetArgIleProArgIleSerAspIleAlaGlyThr 239  
QY 712 ACCAGACCTGTGAGAGAGAGATGCGCTGATGATCAACGACACCCCGCCATCCCGTGGG 771  
Db 240 ThrSerThrLeuGlnGluGlnIleGlyTyrMetCThrAsnAsnProIleProValGly 259  
QY 772 GACATCTACAGAGGCTGATCATCTGGGCTGACACAGATCGTGCGAGATGACAGCCG 831  
Db 260 GluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerPro 279  
QY 832 GTGAGCATCTGACATCAACAGAGGCGCCAGAGACCCCTTCGCGACTACGTGACCG 891  
Db 280 ThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAspArg 299  
QY 892 TTCTTCAAGACCTGTGGCGCGCGACGACGACCCAGAGAGTGAAGAACTGTGATGACCG 951  
Db 300 PheTyrLysThrLeuArgAlaGluGlnAlaSerGlnIleValLysAsnTrpMetTrpGln 319  
QY 952 ACCCTCTGTGTGAGAGAGCGCAACCCCGACTGTGACAGCATCTGCGCGCTGCGCGCC 1011  
Db 320 ThrLeuLeuValGlnMetAlaAsnProPheCysLysThrIleLeuLysAlaLeuGlyPro 339  
QY 1012 GCGCGCAGCTGTGAGAGATGATGATGACCGCTTCGCGAGCGCGCGCGCCAGCCACAG 1071  
Db 340 AlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyIleLys 359  
QY 1072 GCCCGGTGTGTGGCGCGAGCGATGACCGACGACGACCAACGACG-----GTGATGATGCG 1125  
Db 360 AlaArgValLeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGln 379  
QY 1126 AAGAGACATTCAAGGCGCCCGCGCATCTGCAAGTGCTTCAATGCGCGCAAGAGAGG 1185  
Db 380 ArgGlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnGlySerLysGluGly 399  
QY 1186 CACATGCGCGCAACTGCGCGCGCGCGCGCGCAAGAGGCTGTGAAATGCGCGCAAGAG 1245  
Db 400 HisIleAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGln 419  
QY 1246 GGCACCAATGAAGATGACGACCGGCGCGCGCAACTTCTTGGGCAAGATCTGGCC 1305  
Db 420 GlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpPro 439  
QY 1306 AGCGACAAAGGCGCGCGCGCAACTTCTGAGAGGCGCGCGCGCGCAACCGCGCGCCG 1365  
Db 440 SerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProPro 459  
QY 1366 GCCGAGAGCTTCGCG-----TTGAGAGAGACCAACCGCGCGCGCAAGAGAGAGCAAG 1419  
Db 460 GluGlnSerPheArgSerGlyValGluThrThrProSerGlnLysGlnGluProIle 479  
QY 1420 GACCGCGAG-----ACCTGACACGCTGAGAGAGCTGTGGCAACGACCGCTGAGC 1473  
Db 480 AspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsnAspProSerSer 499  
QY 1474 CAG 1476  
Db 500 Gln 500

RESULT 5  
FOVWH4  
gag polyprotein - human immunodeficiency virus type 1 (isolate CDC-451)

N:Alternate names: core polyprotein  
N:Contains: core protein p15; core protein p17; core protein p24  
C:Species: human immunodeficiency virus type 1, HIV-1  
C/Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #ext\_change 09-Jul-2004  
C/Accession: A25523  
R:Desat, S.M.; Kalyanaraman, V.S.; Casey, J.M.; Srinivasan, A.; Andersen, P.R.; Devare, A.; Proc. Natl. Acad. Sci. U.S.A. 83, 8380-8384, 1986  
A>Title: Molecular cloning and primary nucleotide sequence analysis of a distinct human A/reference number: A94136, PMID:87041461, PMID:3490666  
A/Accession: A25523  
A/Molecule type: DNA  
A/Residues: 1-500 <DES>  
A/Cross-references: UNIPROT:P05887; GB:M13136; NID:g326459; PIDN:AAA44306.1; PID:g32646.  
C:Geneticus:  
A/Gene: gag  
C/Superfamily: AIDS-related virus gag polyprotein  
C/Keywords: core protein; polyprotein  
F:1-132/Product: core protein p17 #status predicted <p17>  
F:133-391/Product: core protein p24 #status predicted <p24>  
F:392-500/Product: core protein p15 #status predicted <p15>  
Alignment Scores:  
Pred. No.: 4,27e-107 Length: 500  
Score: 2249.00 Matches: 422  
Percent Similarity: 92.22% Conservative: 40  
Best Local Similarity: 84.23% Mismatches: 29  
Query Match: 81.13% Indels: 10  
DB: 1 Gaps: 5  
US-09-475-704A-3 (1-1479) x FOVWH4 (1-500)  
QY 1 ATGGCGCGCGCGCGCGAGATCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 1 MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspArgTrpGluLysIleArg 20  
QY 61 CTGGCGCGCGCGCGCGCGAGAGTGTCTACATGATGAAGACCTGTGTGGCGCGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysGlnTyrArgLeuLysHisIleValTrpLysAspArgLys 40  
QY 121 CTGAGAAAGTTGCGCTGAAACCCCGCGCTGTGTGAGACCGAGCGCGGTGACAGCATG 180  
Db 41 LeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerLysGlyCysArgGlnIle 60  
QY 181 ATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 61 LeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGlnGluLeuArgSerLeuTyrAsn 80  
QY 241 ACCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 81 ThrValAlaThrLeuTyrCysValHisGlnArgIleGluValArgAspThrLysGluAla 100  
QY 301 CTGAGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 119  
QY 361 GCCGACAAAGGCG-----AAGTGAAGCGCAAGTACCCATGCTGTGAGAACTGTGAG 411  
Db 120 AlaAspThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleValGlnAsnLeuGln 139  
QY 412 GGCAGATGATGACAGAGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 471  
Db 140 GlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValIle 159  
QY 472 GAGAGAAAGCTTCAAGCGCGCGAGGTATCCCATGTTTCAACGCGCTGAGCAAGAGCGCG 531  
Db 160 GluGluLysAlaPheSerProGluValIleProMetPheAlaIleLeuSerGluGlyAla 179  
QY 532 ACCCCCGAGACCTTGAACAGATGTTGAACACCTGTGGCGGCCACCGCGCCATGCGAG 591  
Db 180 ThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGln 199  
QY 592 ATGCTGAAGACACCATCAACAGAGAGCGCGAGTGGAGCCGCTGCACCCCGTGAC 651

```

Db      200 MetLeuLyGluThrIleAenGIuGIuAlaIaGluTrpAspArgLeuHISProValHis 219
QY      652 GCGGGCCCATCGCCCCCGCCGAGAGCCGCGAGCCCCGCGGCGAGCAATCGCCGCGAC 711
Db      220 AAGAGYPrIleAlaProGIuMetArGIuProArGIuSerAspIleAlaGIuThr 239
QY      712 ACCAGACCCCTGAGAGAGAGATCGCTGGATGACCAACCCCCCACTCCCGGAGGC 771
Db      240 ThrSerThleuGIuInIuGlnIleGIuTrpMetThAsnAsnProThrProValGIu 259
QY      772 GACATCTCAAGAGCGTGATCATCTCGGCGCTGAAACAAGATCGTGCGAGTGAACGCC 831
Db      260 GIuIleTyIleArgTrpIleIleLeuGIuLeuAsnIleValArgMetTyIleSerPro 279
QY      832 GTGAGATCTCTGACATCAAGCAGGGCCCCCAAGAGACCTTCGCGACATACGTGAACGC 891
Db      280 IleSerIleLeuAspIleArgTrpGIuGIuProLyGIuProPheArgAspTyIleAlaParg 299
QY      892 TTCTTCAAGACCCCTGCGCCCGAGAGAGCAACCCAGAGGTGAAGATGATGACCGAG 951
Db      300 PheTyIleTyIleThleuArgAlaGIuGIuAlaSerGIuValIleAsnTrpMetThrGIu 319
QY      952 ACCCTGCTGTGTGAGAGAGCCCAACCCCGACTGCAAGACCATCTCGCGCGCTTCGCGCCC 1011
Db      320 ThrLeuLeuValGIuAsnAlaAsnProAspCysIleThrIleLeuIleValGIuGIuPro 339
QY      1012 GCGCGCAGCTGAGAGAGATGATGACCGCGCTGCGAGGGGTGGCGCGCCCGCCAGCAAG 1071
Db      340 AlaAlaThrThleuGIuInIuMetMetTrAlaCysGIuGIuValGIuGIuProGIuAsnIle 359
QY      1072 GCGCGCGTGTGCGCCGAGCGAGTGAAGCCAGGCCCAACACAGC-----GTGATGATGACG 1125
Db      360 AlaArgValIleuAlaGIuIleuIleuSerGIuValIleThAsnSerAlaThrIleuMetThrGIu 379
QY      1126 AAGAGCAATCTCAAGGGCCCCCGCGCATGCTCAAGTGTCTTCACTGCGCGCAAGAGGGC 1185
Db      380 ArgGIuAsnPheArgTrpGIuGIuGIuValIleCysPheAsnIleCysGIuGIuGIuGIuGIu 399
QY      1186 CACATCGCCGCGCACTGCGCGCGCCCGCGCAAGAGGGGTGTGAGAGGGCGCAAGAGG 1245
Db      400 HisIleIleAlaArgAsnCysIleAlaProAlaGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 419
QY      1246 GCGCCACAGATGAAGAGTGAACCGAGCGCCAGGCCCACTTCCTGGGCAAGATCTGAGCC 1305
Db      420 GIuIleIleAlaArgMetIleAspCysThrGIuArgGIuAlaAsnPheLeuGIuIleTrpPro 439
QY      1306 ACCCAACAAGGGCGCGCCCGCAACTTCTGCGAGAGCGCGCGCGAGCCCAACCGCCCGCC 1365
Db      440 SerHisIleGIuGIuArgProGIuAsnPheLeuGIuInIleSerArgProGIuProThrAlaProPro 459
QY      1366 GCGGAGAGCTTCGCGCTTC-----GAGAGACCAACCCCGCGCAGAGAGAGAGCAAG 1419
Db      460 GIuGIuIleSerPheArgPheGIuAspGIuIleuTrpThrTrpProSerGIuIleGIuGIuGIuGIu 479
QY      1420 GACCGCGAG-----ACCCTGACCAAGCTGAAGAGCTGTTCGGACACGACCCCTGAGC 1473
Db      480 AspIleGIuIleuTyIleProIleuAlaSerIleuArgSerIleuPheGIuAsnAspProSerSer 499
QY      1474 CAG 1476
Db      500 Gln 500

```

## RESULT 6

## FOVMA2

gag polyprotein - human immunodeficiency virus type 1 (isolate ARV-2)

N:Alternate names: core polyprotein

M:Contains: core protein p15; core protein p17; core protein p24

C:Species: human immunodeficiency virus type 1, HIV-1

A:Note: host Homo sapiens (man)

C:Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004

C:Accession: A03947

R:Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh

Science 227, 484-492, 1985

```

A:Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).
A:Reference number: A04003; MUID:85090453; PMID:2576227
A:Accession: A03947
A:Molecule type: DNA
A:Residues: 1-502 <S>N>
A:Cross-references: UNIPROT:P03349; GB:K02007; NID:G328658; PIDN:AAB59875.1; PID:G328661
C:Comment: Cleavage sites that yield the mature core proteins remain to be determined.
C:Genetics:
A:Gene: gag
C:Superfamily: AIDS-related virus gag polyprotein
C:Keywords: AIDS; core protein; immunodeficiency; polyprotein
F:1-134/Product: core protein p15 #status predicted <P17>
F:135-393/Product: core protein p24 #status predicted <P24>
F:394-502/Product: core protein p15 #status predicted <P15>

Alignment Scores:
Pred. No.: 5,39e-107 Length: 502
Score: 2247.00 Matches: 418
Percent Similarity: 92.43% Conservative: 46
Best Local Similarity: 83.27% Mismatches: 28
Query Match: 81.06% Indels: 10
DB: 1 Gaps: 4

US-09-475-704a-3 (1-1479) x FOVMA2 (1-502)
QY      1 ATGGCGCGCGCGCGAGCATCTGCGCGCGCGCAAGCTGACGCGCTGGAGAGCATCCGC 60
Db      1 MetGIuAlaArgAlaSerValLeuSerGIuGIuLeuAspIleTrpGIuIleArg 20
QY      61 CTGCGCGCGCGCGCAAGGTGTACTGATGGAAGCACTGTGTGTGGCGCGCGAG 120
Db      21 LeuArgProGIuGIuIleGIuIleGIuIleGIuIleGIuIleGIuIleGIuIleGIuIle 40
QY      121 CTGGAAGAATTGGCCCTGAACCCCGGCTGTGAGAGACACAGAGGGTGTCAAGCATTC 180
Db      41 LeuGIuArgPheAlaValAlaAsnProGIuLeuGIuIleTrpSerGIuGIuGIuGIuGIuGIu 60
QY      181 ATCCGCGAGCTGCACCCCGCTGCAAGACCGCGAGCGAGAGCTGAAGAGCTGTGAAC 240
Db      61 LeuGIuIleuGIuInIleProSerIleuGIuInIleGIuIleGIuIleGIuIleGIuIleGIuIle 80
QY      241 ACCGTGGCACCTGTACTGTGTGCAACGAAGATCGAGTCCGCGACCAACCAAGAGGCC 300
Db      81 ThrValAlaThrIleuTyIleValIleGIuIleArgIleAspValIleAspThrIleGIuIle 100
QY      301 CTGACCAAGATCGAGAGAGAGAGCAAGCAAGTCCAGCAAGAAATCCAGAGGCCGAGGCC 360
Db      101 LeuGIuIleGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 120
QY      361 GCC-----GACCAAGGCAAGGTGAGCCAGAACTACCCCATCGTGCAGAACTCG 408
Db      121 AlaAlaGIuThrGIuIleAsnSerSerGIuValSerGIuIleuTrpProIleValGIuIleAsnLeu 140
QY      409 CAGGCGCAGATGTGACACAGAGCATACGCCCGCGACCTGAACGCTGTGGTGAAGTG 468
Db      141 GIuGIuIleMetValHisGIuAlaIleSerProArgTrpIleuAlaIleTrpValIleVal 160
QY      469 ATCGAGAGAGAGCGCTTCAGCCCGGAGGTATCCCATGTTCACCCGCTGAGAGGAGGCC 528
Db      161 ValGIuGIuIleValIlePheSerProGIuValIleProMetPheSerIleAlaLeuSerGIuGIu 180
QY      529 GCCACCCCGCGAGCTGGAACAGATGTGAACACGCTGGGCGCGCAACGAGCGCGCATG 588
Db      181 AlaThrProGIuIleAsnIleAsnIleMetIleuAsnThrValGIuGIuIleAlaIleMet 200
QY      589 CAGATGTGAAGAGCACCATCAACAGAGAGCGCGCGAGTGAAGCGCGTGCACCCGTG 648
Db      201 GIuMetIleuGIuGIuTrpIleAenGIuGIuAlaIleGIuTrpArgValIleIleProVal 220
QY      649 CAGCCCGCGCGCATCGCGCGCGAGATGCGGAGCGCGCGCGAGATGAGCGCGGC 708
Db      221 HisIleGIuProIleAlaProGIuGIuIleMetArgIuProArgIleSerAspIleAlaGIu 240

```

```

QY 709 ACCACGACGACCTGACGAGACGATGCGCTGGATGACGCAACCCCGCATCCCGT 768
    |||||
Db 241 ThrThrSerThrLeuGlnGlnGlnIleGlyTyrMetThrAsnProProlIleProVal 260
QY 769 GGGGACATCTACAGCGGTGGATCATCTGGGCTGAAACAGATCGTGGATGTACAG 828
    |||||
Db 261 GlyGlnIleTyrIleArgTyrPheIleLeuGlyLeuAsnIleValArgMetTyrSer 280
QY 829 CCCGTAGCATCTTGACATCAAGCGGCGCCAGAGGCCCTTCGCGCATACGTGAC 888
    |||||
Db 281 ProThrSerIleLeuAsnIleArgGlnGlyProIleGlnIleProPheArgAspTyrValAsp 300
QY 889 CGCTTTCACAGCCCTGGCGCGCGGACGACGACCGAGGAGTGAAGATCTGATGAC 948
    |||||
Db 301 ArgPheTyrThrLeuArgIleGlnGlnIleSerGlnAspValIleAsnTyrMetThr 320
QY 949 GACACCTGCTGTGACAGACGCAACCCCGACTGACAGACATCTGCGCTCTCGG 1008
    |||||
Db 321 GlnThrLeuLeuValGlnAsnIleAsnProAspCysIleThrIleLeuIleValGly 340
QY 1009 CCCGCGCGACGCTGACGAGATGATGACCGCTGCCAGGCGCTGGCGCGCCGACGAC 1068
    |||||
Db 341 ProIleAlaThrLeuGlnIleGlnIleMetThrAlaCysGlnGlyValGlyProGlnIle 360
QY 1069 AAGCGCGCGGTGCTGGCGGACGAGGATGACGCGCAAC-----ACGAGCGTATGATG 1122
    |||||
Db 361 LysAlaIleValLeuAlaGlnIleMetSerGlnValThrAsnProAlaAsnIleMet 380
QY 1123 CAGAAAGCACTTCAGAGGCGCCCGCGCATGCTCAAGTGTCTTCACTGCGCAAGAG 1182
    |||||
Db 381 GlnArgGlyAsnAsnThrArgAsnGlnArgGlyThrValIleCysPheAsnIleGly 400
QY 1183 GCGCATGTCGCGCACTGCGCGCGCGCGCGCGCGCAAGAGGCTGTGAAATGCGCAAG 1242
    |||||
Db 401 GlyHisIleAlaIleAsnIleCysAlaIleProArgIleValGlyCysThrArgCysGly 420
QY 1243 GAGGGCGACAGATGAAAGATGACGACGCGCGCGCAAGCTTCTGGGCAAGATCTG 1302
    |||||
Db 421 GlnGlnIleGlnIleMetIleAsnProCysThrArgIleAlaAsnIleAsnIleGly 440
QY 1303 CCCAGCGCAAGGCGCGCGCGCGCACTTCTGACAGCGCGCGCGCGCAAGCGCGCG 1362
    |||||
Db 441 ProSerTyrIleGlyArgProGlnIleAsnIleGlnIleSerArgProIleProThrAlaPro 460
QY 1363 CCCGCGGACGCTTCCGCTT-----GAGAGACCAACCCCGCGCAAGACGAGAC 1416
    |||||
Db 461 ProGlnIleSerThrArgPheGlyGlnIleGlyThrThrProSerGlnIleGlnIlePro 480
QY 1417 AAGGACCGGAG-----ACCTGACACGCTGAAAGCTGTGGCAAGACCGCGT 1470
    |||||
Db 481 IleAspIleGlnIleuTyrProLeuIleThrSerLeuArgSerLeuArgIleAsnProSer 500
QY 1471 AGCCAG 1476
    |||||
Db 501 SerGln 502

```

## RESULT 7

```

FOVWH3
gag polyprotein - human immunodeficiency virus type 1 (isolate HTLV-III, BH10)
N:Alternate names: core polyprotein
M:Contains: core protein p15; core protein p17; core protein p24
C:Species: human immunodeficiency virus type 1, HIV-1
A>Note: host Homo sapiens (man)
C:Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C:Accession: A03945
R:Author: L.; Haseltine, W.; Patarca, R.; Livak, K.J.; Starich, B.; Josephs, S.F.; Dore
nberger, J.A.; Papas, T.S.; Graybe, J.; Chang, N.T.; Gallo, R.C.; Wong-Staal, F.
N:Title: Complete nucleotide sequence of the AIDS virus, HTLV-III.
A:Reference number: A93353; M01D:8511123; PMID:2578615
A:Accession: A03945
A:Molecule type: DNA
A:Residues: 1-512 <RAT>

```

```

A:Cross-references: UNIPROT:P03347; GB:M15654; GB:K02008; GB:K02009; GB:K02010; NID:9326
C:Comment: Cleavage sites that yield the mature core proteins remain to be determined.
C:Genetics:
A:Gene: gag
C:Superfamily: AIDS-related virus gag polyprotein
C:Keywords: AIDS; core protein; immunodeficiency; polyprotein
F:1-112/Product: core protein p17 #status predicted <P17>
F:133-391/Product: core protein p24 #status predicted <P24>
F:392-512/Product: core protein p15 #status predicted <P15>

```

Alignment Scores:

Pred. No.:	2,19e-106	Length:	512
Score:	2235.00	Matches:	424
Percent Similarity:	90.64%	Conservative:	41
Best Local Similarity:	82.65%	Mismatches:	26
Query Match:	80.63%	Indels:	22
DB:	1	Gaps:	7

US-09-475-704a-3 (1-1479) x FOVWH3 (1-512)

```

QY 1 ATGGGCGCGCGCGCGACGATCTTGGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCG 60
    |||||
Db 1 MetGlyAlaArgAlaSerValIleuSerGlyGlnIleuAspArgTyrGlnIleArg 20
QY 61 CTGGCGCGCGCGCGCAAGAGTGTACATGATGAAAGCACTGTGTGGCGCGCGCGAG 120
    |||||
Db 21 LeuArgProGlyValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 40
QY 121 CTGGAGAAATGTCCTGAAACCCCGCGCTGTGAGACCGAGCGAGGCTGCAAGATC 180
    |||||
Db 41 LeuGlnArgPheAlaValAsnProGlyLeuIleuGlnIleuIleuIleuIleuIleuIleu 60
QY 181 ATCCGCGACGCTGACCGCGCGCTGCAAGCGCGCGAGGAGAGCTGAAAGCTTTCAC 240
    |||||
Db 61 LeuGlyGlnIleuGlnIleuProSerIleuGlnIleuGlnIleuIleuIleuIleuIleu 80
QY 241 ACCGCGCGCGCGCTGATCTGATGAGTGAAGATGAGTGGCGCGCGCAAGAGGCG 300
    |||||
Db 81 ThrValAlaThrLeuIleuTyrValIleuGlnIleuIleuIleuIleuIleuIleuIleu 100
QY 301 CTGCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
    |||||
Db 101 LeuAspIleIleGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 119
QY 361 GCGCAAGAGG-----AAGTGACCGCAAGATCAACCCCATGCTGTGCAAGCTGCA 411
    |||||
Db 120 AlaAspThrGlnIleSerSerGlnIleValSerGlnIleValIleGlnIleGln 139
QY 412 GCGCGAGTGTGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 471
    |||||
Db 140 GlyGlnMetValIleGlnIleIleSerProArgTyrIleuAsnIleAlaIleValIleVal 159
QY 472 GAGGAGAGGCGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 531
    |||||
Db 160 GlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 179
QY 532 ACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 591
    |||||
Db 180 ThrProGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 199
QY 592 ATGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 651
    |||||
Db 200 MetLeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 219
QY 652 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
    |||||
Db 220 AlaGlyProIleAlaIleProGlyIleuIleuIleuIleuIleuIleuIleuIleuIleu 239
QY 712 ACCAGACCGCTGAGAGAGAGATGCTGTGATGACGAGCAACCGCGCGCGCGCGCG 771
    |||||
Db 240 ThrSerThrLeuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 259
QY 772 GACATCTACAGCGGTGATCATCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 831

```



```

      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
      260 GLLIETLYLVAHRTGRIETLEULEUGLYLENAHNLVLEVALHRTGCTYSERPRO 279
      832 GTGAGCATCTTGACATCAAGAGGCGCCCAAGAGCCCTTCGCGACTTACGTGACCGC 891
      280 ThrSerILEuNApRIleArgINGlyProLYeGLuProPhenArgpHYrValaPaRG 299
      892 TTCTTCAAAACCTGGCGCCCGAGAGACCCAGAGAGTGAAGAACTGTGATGACCGAC 951
      300 PheTYrLysThrLeuArglaGLuGlnAlaSerGlnGluValLysAntPrMetThrGln 319
      952 ACCCTGCTGTGAGAACCCCAACCCGCACTGCAAGACCATCTGCGGCTCTCGCGCC 1011
      320 ThrLeuValGlnHnAlaenProaPbCYsLYThrILEuLYaHaleuGLYPro 339
      1012 GGGCGGACCTTGAGAGATGATGACCGCTGCGAGGCGTGGCGCGCCCAAGCAAG 1071
      340 AlaAlaThrLeuGluGluMetMetThrAlaCYsGlnGlyValGlyProGlyHISLys 359
      1072 GCGCGGTGCTGGCGAGCGATGAGCCAG---GCCAAGACC---AGCGTATGATGCGAG 1125
      360 AlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnThrAlaThrILEuMetGln 379
      1126 AAGAGCAACTTCAAGGCGCGCGCGCATGTCATGCTTCAACTGCGGCAAGAGGCGC 1185
      380 ArgGLYAsnPhenArgsnGlnArgLYsMetValLYsCYsPhenAsnCYsGLYsGLY 399
      1186 CACATGCGCGCGCACTGCGCGCGCGCGCGCAAGAGGCTGTGAGAGTGGCGCAAGAG 1245
      400 HIsThrAlaArgAsnCYsArgAlaProArgLYsGLYCYsThrLYsCYsGLYsGLY 419
      1246 GCGCAACGATGAGAGCTGACCGCGCGCGCGCGCAACTTCTGCGGCAAGATCGCGCC 1305
      420 GLYHISGlnMetLYsAspCYsThrGlnArgGlnAlaAsnPhenLeuLYsILEuPro 439
      1306 AGCCCAAGAGGCGCGCGCGCGCGCAAC----- 1329
      440 SerTYrLYsGLYArgProGlnAsnPhenLeuGlnSerArgProGlnProThrAlaProPro 459
      1330 TTCCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGAGACTTCGCGC-----TTC 1383
      460 PheLeuGlnSerArgProGlnProThrAlaProGlnGlnSerPhenArgserGLYAla 479
      1384 GAGGAGACCAACCGCGCGCGCGAGAGAGAGAGCAAGAGAGCGGAG-----ACCGTACC 1437
      480 GlnThrThrThrProProGlnLYsGlnGlnProILEAspLYsGlnLeuTYrProLeuThr 499
      QY 1438 AGCTGAAGAGCGCTGTTCGCAACGACCCCTGAGCGCAG 1476
      Db 500 SerLeuArgSerLeuPheGlyAsnAspProSerSerGln 512
  
```

RESULT 8

gag polyprotein - human immunodeficiency virus type 1 (strain JRF1)

C1Species: human immunodeficiency virus type 1, HIV-1

C1Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #ext\_change 09-Jul-2004

C1Accession: T09436

R1Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie, submitted to the EMBL Data Library, July 1996

A1Reference number: Z16673

A1Accession: T09436

A1Status: preliminary; translated from GB/EMBL/DBJ

A1Molecule type: DNA

A1Residues: 1-498 <PAN>

A1Cross-References: UNIPROT:Q75754, EMBL:U63632, NID:g1465777, PID:g1465778

C1Genetics:

A1Gene: gag

C1Superfamily: AIDS-related virus gag polyprotein

Alignment Scores:

Pred. No.: 2,77e-106 Length: 498

Score: 2233.00 Matches: 417

Percent Similarity: 92.57% Conservative: 44

```

      Best Local Similarity: 83.73%
      Query Match: 80.56%
      DB: 2
      Gaps: 4
      US-09-475-704A-3 (1-1479) x T09436 (1-498)
      1 ATGGCGCGCGCGCGCGAGCATCTGCGCGCGCGCGAGAGCTGAGACCTGGAGCGCATCCGC 60
      1 MetGlyAlaArgAlaSerValLeuSerGlyLYsLeuAspLYsTrpGluLYsILEArg 20
      61 CTGCGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGTGTGGCGCGCGCGAG 120
      21 LeuArgProGlyLYsLYsLYsLYsTrArgLeuLYsHISILEValThrLaserArgGlu 40
      121 CTGGAAGAGTTGCGCCTGAACCCGCGCTGCTGAGACCAAGCGAGCGCTGCAAGCAATC 180
      41 LeuGluArgPheAlaValaAsnProGlyLeuLeuGlnSerSerGlnGlyCYsArgGlnILE 60
      181 ATCCGCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGAGCGAGCTGAAGAGCTGTTCAAC 240
      61 LeuGlyGlnLeuLeuProAlaLeuLYsThrGlySerGlnGluLeuArgSerLeuTYrAsn 80
      241 ACCGTGCGCACCTGTACTGTGCGTGAAGAAAGATCGAGTCCGCGCAACCAAGAGCGCC 300
      81 ThrValAlaThrLeuTYrCYsValHISGlnArgILEGluValLYsAspThrLYsGluAla 100
      301 CTGGAAGAGTGAAGAGAGAGCAAGCAAGTGCAGCAAGAAAGATCCAGAGCGCGAGGCC 360
      101 LeuGluLYsILEGluGlnGlnAlaLYsSerLYsLYsGluAlaAlaAlaAspThr 120
      361 GCGCAAGAGGAGAGTGAAGAGAGCAAGCAACCCCATCGGCGAAGCTGCGAGGCGCGATG 420
      121 GlyAsnSerSerGlnValSerGlnAsnTYrProILEValGlnAsnMetGlnGlyMet 140
      421 GTGCAACGAGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCATGAGTGAAGAG 480
      141 ValHISGlnAlaILESerProArgThrLeuAsnAlaTrpValLYsValaGluGluLYs 160
      481 GCGTTGAGCGCGAGAGTATCCCATGTTCAACCGCGCTGAGCGAGGCGCGCGCGCGAG 540
      161 AlaPheSerProGluValaILEProMetPheSerAlaLeuSerGlnGlyAlaThrProGln 180
      541 GACCTGAACGAGATTTGAACACGCTGGCGCGCGCGCGCGCGCGCGCGCATGAGTGAAG 600
      181 AspLeuAsnThrMetLeuAsnThrValaGlyLYsGlnAlaAlaMetGlnMetLeuLYs 200
      601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGACCGCGTGAACCCCGTGCACGCGCGCC 660
      201 GlnThrILEaGlnGluAlaAlaGluTrpAspArgLeuHISProValHISAlaGlyPro 220
      661 ATCGCGCGCGCGCGAGATCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
      221 IleaLaprogLYsGlnMetArgGluProArgLYsSerPheILEaGlyThrThrSerThr 240
      721 CTGGAAGAGAGATGCGCTGATGATGACGAGACCCCGCGCGCGCGCGCGCGCGCATCTAC 780
      241 LeuGlnGlnGlnILEGlyTrpMetThrAsnAsnProProILEProValaGlyGluILETYr 260
      781 AAGCGGTGATCATCTGCGCGCTGAACAGATCGTGGAGTGAACAGCCCGGTGAGCATC 840
      261 LysArgTrpILEILEuGlyLeuAsnLYsILEValaArgMetTYrSerProThrSerILE 280
      841 CTGGAACATCAAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTTCAAG 900
      281 LeuAspILEaGynGlnGlyProLYsGluProPhenArgpHYrValaPaRgPheTYrLYs 300
      901 ACCCTGCGCGCGCGAGAGAGCAACCAAGAGAGTGAAGAACTGATGAGCAAGCAACCTGTG 960
      301 ThrLeuArglaGLuGlnAlaSerGlnGluValLYsAntPrMetThrGlnThrLeuLeu 320
      961 GTGCAAGAGCGCAACCCGCGCTGCAAGACCATCTGCGCGCTGCGCGCGCGCGCGCGC 1020
      321 ValGlnAsnAlaAsnProaPbCYsLYeThrILEuLYsAlaLeuGlyProAlaAlaThr 340
  
```



QY 1021 CTGAGAGATGATGACCGCTGCGAGGGGCGGCCCAAGAGCCCGCTG 1080  
DB 341 LeuGIGlmeTethrAlaCYsGInGlyValGlyProGlyhIshYsAlaGVal 360  
QY 1081 CTGGCCGAGCGATGAGCCAGGCCAAC-----ACCAGCTGATGATGACAGAGAGCAAC 1134  
DB 361 LeuAlAGlulhIamTerSerGlnValThraAnProAlaThrIleMetGlnAraGlyAen 380  
QY 1135 TTCAAGAGGCCCCCGGCGATCTTCAGAGTCTTCACTGCGGCAAGAGGGCCCATCGCC 1194  
DB 381 PheArAsnGlnhIarGlyhIeValIySvPheAsnCYsGlyhIySgInGlyhIleAla 400  
QY 1195 CGCACTGCGCGCGCCCCCGGCAAGAGGGGCTGTGGAAGTGGCGCAAGAGGGCCACCG 1254  
DB 401 ArgAsnCYeArhGAlaProhIarGlyhIySgLYCyserTrpYCyseGlyhIySgInGlyhI 420  
QY 1255 ATGAAGAGTGCACCGAGCGCCAGCCCACTTCCTGGCAAGATCTGCGCCAGCCACAG 1314  
DB 421 MetLyAspCYeArhGInrGlnrGlnAlaAenPheGlnIyYsIleTrpProSerTrlyS 440  
QY 1315 GGCAGCGCCCGGCACTTCTCTGAGAGCGCGCCCGAGCCACCGCCCCCGCGAGAGC 1374  
DB 441 GlyArgProGlyAsnPhelGlnSerArgProGlnProThrAlaProProGlnIyYsSer 460  
QY 1375 TTCCGCTC---GAGAGAGC---ACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAG 1428  
DB 461 PheArhPheGlyGlnIulhIarAlaTrnProSerGlnhIySgInIleArpIySgIn 480  
QY 1429 -----ACCTGACAGCCTGAAGAGCTGTTCGCAAGAGCCCGCTGAGCCAG 1476  
DB 481 MetLyProLeuThrSerLeuArGserLeuPheGlyAsnAspProSerSerGln 498

## RESULT 9

A44001  
gag polyprotein - human immunodeficiency virus type 1 (strain YU-2)  
N:Alternate names: core polyprotein  
M:Contains: core protein p15; core protein p17; core protein p24  
C:Species: human immunodeficiency virus type 1, HIV-1  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C:Accession: A44001  
R:Lit.: Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.  
J. Virol. 66, 6587-6600, 1992  
A:Title: Complete nucleotide sequence, genome organization, and biological properties of  
A:Reference number: A44001; M0ID:93021387; PMID:1404605  
A:Accession: A44001  
A:Molecule type: DNA  
A:Residues: 1-500 <LTY>  
A:Cross-references: UNIPROT:P35962; GB:M93258  
C:Genetics:  
A:Gene: gag  
C:Superfamily: AIDS-related virus gag polyprotein  
C:Keywords: AIDS; core protein; immunodeficiency; polyprotein  
F:1-132/Product: core protein p17 #status predicted <P17>  
F:133-391/Product: core protein p24 #status predicted <P24>  
F:392-500/Product: core protein p15 #status predicted <P15>

## Alignment Scores:

Pred. No.:	3.5e-106	Length:	500
Score:	2231.00	Matches:	420
Percent Similarity:	91.82%	Conservative:	40
Best Local Similarity:	83.83%	Mismatches:	31
Query Match:	80.48%	Indels:	10
DB:	1	Gaps:	5

US-09-475-704A-3 (1-1479) x A44001 (1-500)

QY 1 ATGGGCGCCGCGCGAGCATCTGCGCGGCGCAAGCTGAGCGCTGAGGCGCATCGC 60  
DB 1 MetGlyAlaArGAlaSerValLeuSerAlaGlyGlnLeuAspIyTrpGlnIySleArG 20  
QY 61 CTGGCGCCCGGCGCAAGAGTGTCTAATGATGAGACACTGTGTGTGGCCAGCGCGAG 120

DB 21 LeuArhProGlyGlyIyYsGlnTrpArgLeuIyhIleValIleTrpAlaSerArGln 40  
QY 121 CTGAGAGATTGACTGAACCCCGGCTGTGAGACCAAGAGGCTGCAAGAGATC 180  
DB 41 LeuGlnArhPheAlaValAspProGlyLeuLeuGlnThrSerGlnIyCyArGlnIle 60  
QY 181 ATCCGCGAGCTGACCCCGCTGAGACCGGCGAGAGAGAGTGAAGACCTTCAC 240  
DB 61 LeuGlyhIleGlnProSerLeuGlnTrnGlySerGlnIyLeuArGserLeuTrpAsn 80  
QY 241 ACCGTGACCACTGACTGCTGTCAGAGAAAGATGAGAGTCCGCGACCAAGAGGCG 300  
DB 81 ThrValAlaThrLeuTrpYsValhIseGlnIySleGlnValIyAspThrIySgInAla 100  
QY 301 CTGACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGATTCAGAGCGCGAGCC 360  
DB 101 LeuGlnIySleGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIy 119  
QY 361 GCGGACAGAGGC-----AAGTGAGCCGAAGTACCCTCATGTGCAAGACTGCGAG 411  
DB 120 AlaAspThrGlyAsnSerSerGlnValSerGlnAsnTrpProIleValGlnAsnLeuGln 139  
QY 412 GCGCAGATGTCACACGAGCCATCAGCGCCCGGCAACCTGAGCGCTGGAAGTGATC 471  
DB 140 GlyhIleValhIseGlnAlaIleSerProhIarGlnhIarAlaTrpAllySvalVal 159  
QY 472 GAGAGAGAGCCTTACGCGCCGAGTGATCCCATGTTTACCGCGCTGAGCGAGCGCC 531  
DB 160 GlnIyIyYsAlaIlePheSerProGlnIyIlePromePheSerAlaLeuSerGlnIyAla 179  
QY 532 ACCCGCCAGAGACTGAACAGATGTTGAACACCGTGGGCGCCACAGGCGCGCATGCGAG 591  
DB 180 ThrProGlnAspLeuAsnThrMetLeuAsnTrpValGlyIyhIseGlnIleAlaIleMetGln 199  
QY 592 ATGCTGAAGAGACACCATCAAGAGAGAGCGCGCGAGTGGGACCGGCTGACCCCGCGAGC 651  
DB 200 MetLeuIySgInTrhIleAsnGlnIyGlnAlaIyGlnIyTrpAspArgLeuIleProValhIAs 219  
QY 652 GCGGCGCCCGATCGCCCGGCGAGTGGCGAGCGCGCGCGGCGAGAGATCGCGCGAGCC 711  
DB 220 AlaGlyProIleAlaIleProGlyGlnMetArGlnProhIarGlySerAlaIleAlaGlyThr 239  
QY 712 ACCAGACCTTGCAGAGAGAGATCGCTGATGATGACCAAGACACCCCGCATCCCGTGGGC 771  
DB 240 ThrSerTrnLeuGlnIyGlnIleGlyTrpMetTrnAsnAspProIleProValGly 259  
QY 772 GACATCTAAGAGCGGTGATCATCTGGGCGCTGAACAAGATGTCGAGTGAACGCGCC 831  
DB 260 GlnIleTrpYsArGTrpIleIleLeuGlyLeuAsnIySleValAlaArgMetYsSerPro 279  
QY 832 GTGACATCTTGAATCAAGAGAGAGCGCCCAAGAGACCTTCGCGACTACGTGAGCCGC 891  
DB 280 ThrSerIleLeuAspIleArgGlnIyProIySgInProPheArhAspTrpValAspArg 299  
QY 892 TTCTTCAAGACCTTCCGCGCGCGAGAGAGACCCAGAGGTGAAGAACTGATGACCGAGC 951  
DB 300 PheTrpIySleTrnLeuArGAlaGlnIyGlnAlaSerGlnIyValIyAsnTrpMetThrGln 319  
QY 952 ACCGTGTCGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
DB 320 ThrLeuLeuValGlnAsnAlaAsnProAspCYsIyTrnIleLeuIyAlaLeuGlyPro 339  
QY 1012 GCGCGCAGCTGAGAGAGATGATGACCGCTGTCAGAGGCGTGGGCGCGCCAGCAAG 1071  
DB 340 AlaAlaTrnLeuGlnIyGlnMetMetThrAlaCYsGlnIyValGlyIyProGlyhIyS 359  
QY 1072 GCGCGGTGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1125  
DB 360 AlaArGValLeuAlaGlnAlaIleMetSerGlnValThraAsnSerAlaThrIleMetMetGln 379  
QY 1126 AAGAGCACTTCAAGAGGCGCGCGCGCATCTGCAAGTGTCTTCAACTGGCGAGAGAGGCG 1185



Db 460 GIUGLUser:PheargPheglYgluGluThrThrProSerGlnlysgInluProile 479  
QY 1420 GACCGGAG-----ACCTGACCAAGCTTGGGCAAGACCCCTGAGC 1473  
Db 480 AApLySGlMeTyTProLeuAlaSerLeuArgSerLeuPheGlyAAsnAspProSerSer 499  
QY 1474 CAG 1476  
Db 500 Gln 500  
RESULT 11  
A38068  
gag polyprotein - human immunodeficiency virus type 1 (strain MN)  
N:Alternate names: core polyprotein  
N:Contains: core protein p1, core protein p17, core protein p2, core protein p24, core F  
C:Species: human immunodeficiency virus type 1, HIV-1  
A:Note: host Homo sapiens (man)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C:Accession: A38068  
R:Henderson, L.B.; Bowers, M.A.; Sowder II, R.C.; Serabyn, S.A.; Johnson, D.G.; Bees Jr.  
J.Virol. 66, 1856-1865, 1992  
A:Title: Gag proteins of the highly replicative MN strain of human immunodeficiency viru  
A:Reference number: A38068; MUID:92194415; PMID:1548743  
A:Accession: A38068  
A:Molecule type: protein  
A:Residues: 1-506 <HEN>  
A:Cross-references: UNIPROT:P05888  
C:Gene(s):  
A:Gene family: AIDS-related virus gag polyprotein  
C:Keywords: AIDS; core protein; immunodeficiency; polyprotein  
F:1-134/Product: core protein p17 #status experimental <P17>  
F:135-365/Product: core protein p24 #status experimental <P24>  
F:366-375/Product: core protein p2 #status experimental <P2>  
F:386-434/Product: core protein p7 #status experimental <P7>  
F:435-456/Product: core protein p1 #status experimental <P1>  
F:451-506/Product: core protein p6 #status experimental <P6>  
Alignment Scores:  
Pred. No.: 1,91e-105 Length: 506  
Score: 2216.50 Matches: 418  
Percent Similarity: 90.34% Conservative: 40  
Best Local Similarity: 82.45% Mismatches: 32  
Query Match: 79.96% Indels: 17  
DB: 1 Gaps: 5  
US-09-475-704a-3 (1-1479) x A38068 (1-506)  
QY 4 GGGCGCGCGCGCAGATCCTGCGCGCGCAAGCTGAGCCTGGAGCGCATCCGCGT 63  
Db 1 GYAlaAlaArgAlaSerValLeuSerGlyGluLeuAspArgTrpGluLysIleArgLeu 20  
QY 64 CGCGCGCGCGCAAGAGTGTACATGATGAGCACTGTGTGGCGCAGCGAGCTG 123  
Db 21 ArgProGlyGlyLysLysLeuTyTyTySerLeuLysValValTrrAlaSerArgGluLeu 40  
QY 124 GAGAAGTTGCGCTGAACCCCGCGCTGTGAGACCAAGCGAGCGCTGCAAGATGATC 183  
Db 41 GluArgPheAlaIleAlaAsnProGlyLeuLeuGluThrSerGlyGlyCysArgGlnIleLeu 60  
QY 144 CGCGAGCTGCAACCGCGCTGCAAGCGCGCAGCGAGAGCTGAAGAGCTGTCAACACC 243  
Db 61 GlynLeuGlnProSerLeuGlnThrGlySerGluGluArgLysSerLeuTyTrpAsnThr 80  
QY 244 GTGGCCACCTTACTGCTGTGACGAGAGATGAGATCGATCCGACACCAAGAGCGCTG 303  
Db 81 ValAlaThrIleuTyTyCysValIleGlnLysIleLysIleLysAspThrIysGlnAlaLeu 100  
QY 304 GACAAGATGAGAGAGAGCAAAAGTCCACGACAGAGATCCAGAGCGCCGAGCGCC 363  
Db 101 GluTyIleGluGluGlnGlnAsnLysSerTyLysLysAlaGlnGlnAla--AlaAla 119  
QY 364 GACAAAGGCG-----AGGTGAGCGCAAGACTACCCCATCTGTGACGAAC 405

Db 120 AspThrGlyAsnArgGlyAsnSerSerGlnValSerGlnAsnTyTProIleValGlnAsn 139  
QY 406 CTGCAAGGCGCAGATGGTGCACCAAGGCATACGCCCCCGACCTGAAACCTCGGTGAAG 465  
Db 140 IleGlnGlyGlnMetValIleGlnAlaIleSerProArgThrLeuAsnAlaTrpValLys 159  
QY 466 GTGATCGAGGAGAAAGGCGCTTACGCGCGAGTATCCCATGTTACACGCGCTGAGCGAG 525  
Db 160 ValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGln 179  
QY 526 GCGCGCACCGCCCAAGAGCCTGAACACGATGTTGAACACCGTGGGCGCCACCAAGCGCGC 585  
Db 180 GYAlaThrProGlnAspLeuAsnThrIleLeuAsnThrValGlyGlyIleGlnAlaIle 199  
QY 586 ATGCAAGATGCTGAGAGACCACTCAACAGAGAGCGCCGAGTGGAGCCGCTGACACCC 645  
Db 200 MetGlnMetLeuTyGluThrIleAsnGluGlnAlaIleGluTrpAspArgLeuHisPro 219  
QY 646 GTGCAAGCGCGCGCGCATGGCGCGCGAGATGCGCGAGCGCGCGGAGCGACATCGCC 705  
Db 220 AlaHisAlaGlyProIleAlaProGlyGlnMetCArgLupProArgGlySerAspIleAla 239  
QY 706 GGCACCAACAGACCTGTCAGAGAGAGATCGCGCTGATGACCAAGACCCCGCATCCCG 765  
Db 240 GlyThrThrSerThrLeuGlnGlnIleGlyTrrPmetThrAsnAsnProIlePro 259  
QY 766 GTGGCGCATCTTCAAGCGGTGATCATCTCGGCGCTGAAACAAGATCGCGATGAC 825  
Db 260 ValGlyGluIleTyTyLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTy 279  
QY 826 AGCGCGTGAGATCCTGACATCAAGAGCGCGCCCAAGAGCGCTTCGCGACATACGG 885  
Db 280 SerProSerSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyVal 299  
QY 886 GACCGCTTCTTCAAGACCTGCGCGCGCGCAGACAGACACCAAGAGAGTGAAGATCGAGT 945  
Db 300 AspArgPheTyTyLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMet 319  
QY 946 ACCGACACCTGCTGTGTGAGAGAGCGCAACCCCGCATGTGACATCTGCGCGCTTC 1005  
Db 320 ThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleuLysAlaLeu 339  
QY 1006 GCGCGCGCGCGCGAGCTGAGAGATGATGACCGCTGCGAGGCGCGCGCGCCACAC 1065  
Db 340 GlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyLysProGly 359  
QY 1066 CACAAGGCGCGGCTGTGCGCGAGCGAGTACCGACCGCAACACAGC-----GTGAT 1119  
Db 360 HisLysAlaArgValLeuAlaGlnAlaMetSerGlnValThrAsnSerAlaThrIleMet 379  
QY 1120 ATGCAAGAGCAACTTCAAGGCGCGCGCGCATGTCAAGTGTTCATCTGCGCGCAG 1179  
Db 380 MetGlnArgGlyAsnPheArgAsnGlnArgLysIleIleLysCysPheAsnCysGlyLys 399  
QY 1180 GAGGCGCATCTGCGCGCGCACTGCGCGCGCGCGCGCGCAAGAGGCTGTGAGAGTGGCG 1239  
Db 400 GlynGlyHisIleAlaLysAsnCysArgAlaProArgLysArgGlyCysTrpLysCysGly 419  
QY 1240 AAGAGGCGCAACAGATGAAGATGACCGAGCGCGCGCAAGCTTCTGCGCGCAAGATC 1299  
Db 420 LysGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIle 439  
QY 1300 TGGCCACCGCAAGAGGCGCGCGCGCAACTTCTGCAAGCGCGCGCGAGCCACCGC 1359  
Db 440 TrpProSerCysLysGlyArgProGlyAsnPheProGlnSerArgThrGluProThrAla 459  
QY 1360 CCGCGCGCGAGAGCTTCGCTTC-----GAGAGACCAACCCCGCGCGCAAGAGAGAG 1413  
Db 460 ProProGluGluSerPheArgPheGlyGluGluThrThrTrpProTyGlnLysGlnGlu 479  
QY 1414 -----AGCAAGACCGCGAGACCTTGACCGAGCTGAAGAGCGCTGCTTC 1455

Db	480	LyylLyGInGInIurhriIleApblyApbLeuTYrProLeuAlaSerLeuLYSerLeuPhe	499
Qy	1456	GGCAAGACCCCCCTGAGCCAG	1476
Db	500	GlyAsnApbProLeuSerGln	506

RESULT 12

## FOLJSI

gag polyprotein - simian immunodeficiency virus SIVcpz

N/Alternate names: core polyprotein

C/Species: simian immunodeficiency virus SIVcpz

A/Note: host Pan troglodytes (chimpanzee)

C/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004

C/Accession: S09983

R/Huet, T.; Cheynier, R.; Meyershan, A.; Roelants, G.; Wain-Hobson, S.

Nature 345, 356-359, 1990

A/Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.

A/Reference number: S09983; MUID:90259077; PMID:2188136

A/Accession: S09983

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-508 &lt;HUE&gt;

A/Cross-references: UNIPROT:P17282; EMBL:X52154; NID:958866; PIDN:CAA36401.1; PID:958867

C/Genetic:

A/Suprafamily: AIDS-related virus gag polyprotein

C/Keywords: AIDS; core protein; immunodeficiency; polyprotein

## Alignment Scores:

Pred. No.:	6, 87e-94	Length:	508
Score:	1989.00	Matches:	365
Percent Similarity:	83.73%	Conservative:	62
Best Local Similarity:	71.57%	Mismatches:	63
Query Match:	71.75%	Indels:	20
DB:	1	Gaps:	4

US-09-475-704a-3 (1-1479) x FOLJSI (1-508)

```

QY 1 ATGGGGGGGGGGGGGCGCATCTCGCGGGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60
   |||||
DB 1 MetGlyAlaArgAlaSerValLeuThrGlyGlyLeuAspArgTrpGluValArg 20
QY 61 CTGGCGCGCGCGCGGCAAGAGTCTACATGATGAACACCTGTGTGGCGCGCGAG 120
   |||||
DB 21 LeuArgProGlyGlyArgGlyArgGlyTrpMetLeuHisLeuValTrpAlaSerArgGln 40
QY 121 CTGGAGAAAGTTGGCTCTGAACCCCGGCTGTGTGGAGAGCCAGAGGGCTGGCAAGATC 180
   |||||
DB 41 LeuGluArgPheAlaCysAspProGlyLeuMetGluSerIleGlyCysThrIleu 60
QY 181 ATCCCGACAGTGCACCCCGCTGTGCAGACCGCGAGAGAGAGCTGAAGAGCTGTTCAAC 240
   |||||
DB 61 LeuGlnGlnLeuGlnProAlaLeuGlyTrpGlySerGlyGlyLeuArgSerLeuPheAsn 80
QY 241 ACCGTGGCCACCTGTACAGCGGTGCGAGAGAGATGAGGTCCGCGACCAAGAGAGCC 300
   |||||
DB 81 ThrLeuAlaValLeuTrpCysIleHisSerAspIleThrValGlnAspThrGlnValAla 100
QY 301 CTGGACAGATCGAGAGAGAGAGAGAGCAAGATGCCAGAGAGATCCAGAGCGCGAGCC 360
   |||||
DB 101 LeuGlnGlnLeuLysArgHisIleGlyGlnGlnSerIleThrGlnSerIleAsnSerGly 120
QY 361 GCCGACAGAGCGCAAGGTGAGCCAG-----AACTAC 390
   |||||
DB 121 SerArgGlnGlyGlyAlaSerGlnGlyAlaSerAlaSerAlaGlyIleSerGlyAsnTrp 140
QY 391 CCATGTCAGAGAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 450
   |||||
DB 141 ProLeuValGlnHisMetAlaGlnGlyGlnMetValHisGlnAlaIleSerProHisGlnTrp 160
QY 451 AACGCTGGGTGAAGGTGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 510
   |||||
DB 161 AsnAlaTrpValIleValValGlnGlnGlnValAlaPheSerProIleValIleProMetPhe 180
QY 511 ACCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570
   |||||
DB 181 SerAlaLeuSerIleGlyAlaLeuProGlnAspValAlaAsnThrMetLeuAsnAlaValGly 200
QY 571 GGCACACAGAGCGCGCATGAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630
   |||||

```

```

DB 201 GlyHisGlnGlyAlaMetGlnValLeuGlyGlnValIleAsnGlnGlnAlaIleGluTrp 220
QY 631 GACCGCTACACCCCGTGCACCGCGGCGCATATCGCCCGCGCAATGCGCGAGCCCGC 690
   |||||
DB 221 AspArgLeuHisIleProThrHisIleArgProIleAlaIleProGlyGlnLeuArgGluProArg 240
QY 691 GGCAGAGAGATGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 750
   |||||
DB 241 GlySerAspIleAlaGlyThrThrSerThrLeuGlnGlnGlnIleGlyTrpThrTrpAla 260
QY 751 AACCCCGCATCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 810
   |||||
DB 261 AsnProProIleProValGlyAspValIleArgArgTrpValIleLeuGlyLeuAsnLys 280
QY 811 ATCGTGGAGATGACAGCCCGTGAACATCTGAGATCATCAAGAGAGAGAGAGAGAGAGAG 870
   |||||
DB 281 ValValArgMetCysAspProValSerIleLeuAspIleArgGlnGlyProLysGluPro 300
QY 871 TTCGCGAGTACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 930
   |||||
DB 301 PheArgAspTrpValAspArgPheTrpLysTrpLeuArgAlaGlnGlnAlaSerGlnGlu 320
QY 931 GTGAAGAACTGATGACACGACACCTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 990
   |||||
DB 321 ValLysAsnTrpMetThrAspThrLeuLeuValGlnAsnAlaAsnProAspCysLysGln 340
QY 991 ATCTGCGCGCTCTCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1050
   |||||
DB 341 IleLeuLysAlaLeuGlyProGlyAlaThrLeuGlnGlnGlnMetMetThrLysGlnGly 360
QY 1051 GTGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1098
   |||||
DB 361 ValGlyGlyProSerHisLeuValAlaArgValLeuAlaGlnAlaMetSerMetValGlnAsn 380
QY 1099 CAGGCAACACAGAGCGTGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1158
   |||||
DB 381 GlnGlyArgAlaAspValPhePheGlnLysGlyGlnGlyAlaGlyProLysArgLysIle 400
QY 1159 AAGTCTTCACTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1218
   |||||
DB 401 LysCysPheAsnGlyGlyGlyGlnGlyIleAlaArgAlaGlnCysLeuValAlaProArgArg 420
QY 1219 AAGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1278
   |||||
DB 421 LysGlyCysTrpArgCysGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 440
QY 1279 GCCAATCTCTGGGAGAGATCTGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1338
   |||||
DB 441 ValAsnPheLeuGlyLysGlyTrpProSerArgSerGlyArgProGlyAsnPheValGln 460
QY 1339 AGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1398
   |||||
DB 461 AsnArgTrpGlnProThrAlaProIleGlnSerTrpGlyTrpGlnGlnGlnGlnGlnGln 479
QY 1399 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
   |||||
DB 480 ---LysSerGlnGlnLysLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 498
QY 1447 AGCCTGTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
   |||||
DB 499 SerLeuPheGlySerAspProSerSerGln 508

```

## RESULT 14

S08435

gag polyprotein - human immunodeficiency virus type 2 D205

C/Species: human immunodeficiency virus type 2 D205, HIV-2 D205

C/Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004

C/Accession: S08435

R/Dietrich, U.; Adamski, M.; Kreutz, R.; Seipp, A.; Kuehnle, H.; Ruebsamen-Waigmann, H.

Nature 342, 948-950, 1989

A/Title: A highly divergent HIV-2-related isolate.

A/Reference number: S08434; MUID:90081881; PMID:2594088

A/Accession: S08435

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-521 <DIE>

A:Cross-references: UNIPROT:P15832; EMBL:X16109

A>Note: this sequence was submitted to the EMBL Data Library, Aug-1989

C:Genetics:

A:Gene: gag

C:Superfamily: AIDS-related virus gag polypeptide

C:Keywords: polypeptide

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
5,7e-64	521	1400.00	266	75	108	70	15
Percent Similarity:		66.98%					
Best Local Similarity:		53.06%					
Query Match:		50.51%					

US-09-475-704A-3 (1-1479) x S08435 (1-521)

```

QY 1 ATGGGCGCGCGCGCGCGCATCTGCGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCG 60
Db 1 MetGlyAlaArgGlyValLeuSerGlyLysThrAspGluLeuGluValArg 20
QY 61 CTGGCGCGCGCGCGCGCAAGAGTCTACATGATGAGACCTGATGGCGCGCGCGAG 120
Db 21 LeuArgProGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 121 CTGGAGAAAGTTCGCGCTGAAACCCCGCGCTGCGAGAGAGAGAGAGAGAGAGAGATC 180
Db 41 LeuAspArgPheGlyLeuValGluSerLeuLeuGluSerLysGlyGlyGlyGlyGly 60
QY 181 ATCCGCGAGCTGACACCCCGCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 61 LeuLysValLeuAlaProLeuValProThrGlySerGluAsnLeuLysSerLeuPheAsn 80
QY 241 ACCGTGGCGCGCGCTGACTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 81 IleValCysValIlePheCysLeuHisAlaGluGluValLysAspThrGluGluAla 100
QY 301 CTGACAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 357
Db 101 -----LysLysIleAlaGluHisLeu 108
QY 358 GCCGCGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 384
Db 109 AlaAlaAspThrGluLysMetProAlaThrAsnLysProThrAlaProProSerGlyGly 128
QY 385 AACTACCCCATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 444
Db 129 AsnTyrPro--ValGlnIleuAlaGlyAsnTyrValHisLeuProLeuSerProArg 147
QY 445 ACCCTGAAGAGCTGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 504
Db 148 ThrLeuAsnAlaTyrValLysLeuValGluGluLysLysPheGlyAlaGluValAlaPro 167
QY 505 ATGTTACACGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 564
Db 168 GlyPheGlnAlaLeuSerGluLysCysThrProTyrAspIleAsnGlnMetLeuAsnCys 187
QY 565 GTGGCGCGCGCGCGCGCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 624
Db 188 ValGlyLysHisGlnAlaAlaMetGlnIleLeuArgGluIleIleAsnGluGluAlaAla 207
QY 625 GAGTGGAGACCGGCTGACACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 684
Db 208 AspTyrAspGlnGlnHisPro--SerProGlyProMetProAlaGlyGlnMetLysAsp 226
QY 685 CCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 744
Db 227 ProArgGlySerAspIleAlaGlyThrThrSerThrValGluGlnIleGlnTyrPhe 246
QY 745 ---ACCAGCAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 801

```

```

Db 247 TyrArgAlaGlnAsnProValProValGlyLysHisLysThrArgArgTyrIleGlnLeu 266
QY 802 CTGAACAAAGATCGTCGAGATGACAGCCCGGTGAGCATCTGACATCAAGAGAGAGAGAG 861
Db 267 LeuGlnLysCysValArgMetTyrAsnProThrAsnIleLeuAspIleLysGlnGlyPro 286
QY 862 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921
Db 287 LysGluProPheGlnSerTyrValAspArgPheTyrLysSerLeuArgAlaGlnIleThr 306
QY 922 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 981
Db 307 AspProAlaValLysAsnThrPheMetThrGlnThrLeuLeuIleGlnAsnAlaAsnProAsp 326
QY 982 TGCAAGACCATCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1041
Db 327 CysLysLeuValLeuLysGlyLeuGlyMetAsnProThrLeuGluGluMetLeuThrAla 346
QY 1042 TGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1101
Db 347 CysGlnGlyIleGlyGlyProGlyGlnLysAlaArgLeuMetAlaGluAlaLeuLysGlu 366
QY 1102 GCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1143
Db 367 AlaLeuThrProAlaProIleProPheAlaAlaValGlnGlnLysAla-----Gly 383
QY 1144 CCGCGCGCGCATCTCAAGCTTCACTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1203
Db 384 LysArgGlyThrValThrCysThrPheAsnGlyLysGlnGlnIleThrAlaArgGlnCys 403
QY 1204 GCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1263
Db 404 ArgAlaProArgArgGlnGlyCysThrPheCysGlyLysThrGlyHisIleMetSerLys 423
QY 1264 TGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1323
Db 424 CysProGluArgGlnAlaGlyPheLeuGlyLysLysLysLysLysLysLysLysLysLys 441
QY 1324 GGCACCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1362
Db 442 ArgAsnProMetThrGlnValProGlnIleValThrProSerAlaProProMetAsn 461
QY 1363 CCGCGCGAGAGCTTCCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1401
Db 462 ProAlaGluGlyMetThrProArgGlyAlaThrProSerAlaProProAlaAspProAla 481
QY 1402 -----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
Db 482 ValGluMetLeuLysSerTyrMetGlnMetGlyArgGlnGlnArgGlnLysSerArgGluArg 501
QY 1426 -----GAGACCTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464
Db 502 ProTyrLysGluValThrGluAspLeuHisLeuAsnSerLeuPheGlyGluAsp 520

```

#### RESULT 15

gag polypeptide - human immunodeficiency virus type 2

C:Species: human immunodeficiency virus type 2, HIV-2

C:Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004

C:Accession: S53091

R:Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.

submitted to the EMBL Data Library, March 1995

A:Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu

A:Reference number: S53091

A:Accession: S53091

A:Molecule type: DNA

A:Status: Preliminary

A:Residues: 1-521 <BNC>

A:Cross-references: UNIPROT:Q76621; EMBL:Z48731; NID:g732718; PIDN:CAA86621.1; PID:g7327

C:Superfamily: AIDS-related virus gag polypeptide

C:Keywords: polypeptide

## Alignment Scores:

Pred. No.:	6,04e-64	Length:	521
Score:	1399.50	Matches:	285
Percent Similarity:	68.75%	Conservative:	78
Best Local Similarity:	53.98%	Mismatches:	119
Query Match:	50.49%	Indels:	46
	2	Gaps:	13

US-09-475-704A-3 (1-1479) x SS3091 (1-521)

```

QY 1 ATGGGCGCCGCGCAGCATCTGCGCGCGGCGAGCTGTGACGCTGTGAGCGCATCCG 60
DB 1 MetGlyAlaArgSerSerValLeuArgGlyLeuValAlaArgGluLeuGluValArg 20
QY 61 CTGGCGCCCGCGCGCAGAGAGTCTACATGATGAAGACCTGTGTGGCGCCAGCGG 120
DB 21 LeuArgProGlyGlyLeuValSerValArgLeuValSerValLeuValAlaValArg 40
QY 121 CTGGAAGAGTTCGCGCTGGAACCGCGCGCTGTGAGACGAGCGGCGTGTGAGAGATC 180
DB 41 LeuArgPheGlyLeuValArgLeuValSerLeuValSerLeuValSerLeuValSer 60
QY 181 ATCCGCGACCTGACCCCGCGCGCTGTGAGACGAGCGGCGAGAGCTGTGAGAGCTGTTCAC 240
DB 61 LeuThrValLeuArgProLeuValProThrGlySerGluValSerLeuValSerLeuPheVal 80
QY 241 ACCGTGTGACCCCTGTACTGCTGTGCGAGAGAGATGAGTCCGCGACACCGAGAGGCC 300
DB 81 ThrValCysValIlePheCysIleIleValArgGluValValLeuValSerThrGluVal 100
QY 301 CTGCAACAATGCGAGAGG-----AGCAAGAACAATGCGAGAGAGAGAGAGAGAGAG 343
DB 101 Lys-LysPheAlaArgArgHleuValAlaGluThrGlyThrAlaGluValMetProVal 120
QY 344 -TCCAGAGCGCGCGAGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402
DB 120 nThrSerArgProThrAlaProProSerGlyArg-GlyGlyAsnThrPro--ValGln 138
QY 403 AACCTGTGAGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 462
DB 139 AsnIleGlyIleAsnThrThrAlaValProLeuSerProArgThrLeuValSerAlaThrVal 158
QY 463 AAGGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 522
DB 159 LysLeuValGluGluValLysPheGlyAlaGluValAlaProGlyPheGlnAlaLeuSer 178
QY 523 GAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582
DB 179 GluGlyCysThrProThrAspIleAsnGlnMetLeuValCysValGlyAspIleGlnAla 198
QY 583 GGCATGCAATGCTGTGAGAGACATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 642
DB 199 AlaMetGlnIleIleAlaArgIleIleAlaGlnGluAlaAlaAspTrpAspAlaGlnHis 218
QY 643 CCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 702
DB 219 ProIle--ProGlyProLeuProAlaGlyGlnLeuArgGlnProArgIleSerAspIle 237
QY 703 GCGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 759
DB 238 AlaGlyThrThrSerThrValAspGluGlnIleGlnTrpMetPheArgAlaGlnAsnPro 257
QY 760 ATCCCGGTGGGCGACATCTACAGAGGCGGTGATCATCTGGCGCTGTGAGAGAGAGAG 819
DB 258 ValProValGlyAsnIleThrArgArgTrpIleGlnIleGlyLeuGlnLysCysValArg 277
QY 820 ATGTACAGCGCGGTGATCTGTGAGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 879
DB 278 MetTrpAsnProThrAsnIleLeuAspValLysGlnGlyProLysGlnProPheGlnSer 297
QY 880 TACGTGACCGCGCTTCTTCAAGACCTGTGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAG 939
DB 298 TyrValAlaArgPheThrLysSerLeuArgAlaGluGlnThrAspAlaValLysAsn 317

```

```

QY 940 TGTATGACCGACACCTCTGTGTGTGCAAGAGCCCAACCCCGAGTGTGCAAGACATCTGCCG 999
DB 318 TrpMetThrGlnThrLeuLeuIleGlnAsnAlaAsnProAspCysLeuValLeuValLys 337
QY 1000 GCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1059
DB 338 GlyLeuGlyMetAsnProThrLeuGluGlnMetLeuThrIleCysGlnIleValGlyGly 357
QY 1060 CCCAGCGACAGAGCGCGCGCTGTGCGCGAGCGAGATGAGCGGCGCAACAGCGGTGATG 1119
DB 358 ProGlyGlnLysAlaArgLeuMetAlaGlnAlaLeuValGluAlaMetThrProAlaPro 377
QY 1120 ATGCAAGAGAGCACTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1179
DB 378 IleProPheAlaAlaAlaGln--GlnArgArgThrIleLysCysThrPheAsnCysGlyLys 396
QY 1180 GAGGCGACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1239
DB 397 GluGlyLysSerAlaArgGlnCysArgAlaProArgArgGlnGlyCysThrLysCysGly 416
QY 1240 AAGAGGCGCGACAGATGAGAGACTGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCG 1299
DB 417 LysThrGlyHisIleMetAlaLysCysProAspArgGlnAlaGlyPheLeuGly--Ile 435
QY 1300 TGGCCGACCGACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1350
DB 436 GlyProThrGlyLys--LysProArgAsnPheProValAlaGlnValProGlnGlyLeu 454
QY 1351 ---CCGACGCGCGCGCGCGCGCGAG-----AGCTTCGCGTGTGAGAGAGAGAGAGAG 1398
DB 455 ThrProThrAlaProProAlaAspProAlaValAlaSerLeuGluLysThrMetGlnGln 474
QY 1399 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
DB 475 GlyArgLysGlnArgGluGlnArgGluArgProThrLysGluValThrGluAspLeuVal 494
QY 1426 -----GAGACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1466
DB 495 HisLeuGluGlnGlyLysThrProCysLysGluThrThrGluAspLeuValSerAsn 514
QY 1447 AGCGTGTTCGCGCAAGAG 1464
DB 515 SerLeuPheGlyThrAsp 520

```

Search completed: March 11, 2005, 15:43:00  
 Job time : 87.7254 secs

**THIS PAGE BLANK (USPTO)**



## Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

**When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 11, 2005, 14:56:45 ; Search time 254.42 Seconds  
(without alignments)  
5953.668 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 2772  
Sequence: 1 atggcgccgcgcgcagcatc.....acgaccctctgagccagtaa 1479

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues  
Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool.p/US09475704/runcat.10032005.140221.14821/app\_query.fasta\_1.3342  
-DB=UniProt.03 -QFMT=fastaan -SUFMT=n2p.rup -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR.SCORER=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=tbl -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09475704\_@CGN.1.1.614\_@runcat.10032005.140221.14821 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG.SCORER=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2606	94.0	492	2	Q9WF75 human immun
2	2599	93.8	492	2	Q9WF69 human immun
3	2512	90.6	492	2	Q9WF66 human immun
4	2512	90.6	492	2	Q9WF72 human immun
5	2459	88.7	492	2	Q8UT62 human immun
6	2455.5	88.6	493	2	Q8AFIG human immun
7	2448	88.3	492	2	Q901X9 human immun
8	2437	87.9	492	2	Q8UT89 human immun
9	2437	87.9	492	2	Q901X0 human immun
10	2435	87.8	492	2	Q8AFW4 human immun
11	2432.5	87.8	499	2	Q8UT79 human immun
12	2432	87.7	492	2	Q8UT71 human immun
13	2429	87.6	492	2	Q8UT53 human immun
14	2427.5	87.4	491	2	Q7SLX7 human immun
15	2424	87.4	492	2	Q7SLX3 human immun
16	2423	87.4	492	2	Q90127 human immun

17	2423	87.4	492	2	Q90CY6 human immun
18	2422.5	87.4	491	2	Q7SLM7 human immun
19	2421.5	87.4	491	2	Q90MFO human immun
20	2421.5	87.4	491	2	Q6X4R8 human immun
21	2420.5	87.3	491	2	Q7SLY0 human immun
22	2420.5	87.3	491	2	Q7SLY1 human immun
23	2420	87.3	492	2	Q667X0 human immun
24	2420	87.3	500	2	Q8AMM2 human immun
25	2420	87.3	500	2	Q8AMM8 human immun
26	2419.5	87.3	491	2	Q90090 human immun
27	2419.5	87.3	501	2	Q8UTP2 human immun
28	2418.5	87.2	853	2	Q8JEH5 human immun
29	2418.5	87.2	873	2	Q8JEG2 human immun
30	2418	87.2	488	2	Q6PR24 human immun
31	2417	87.2	492	2	Q6S859 human immun
32	2416	87.2	492	2	Q8UT71 human immun
33	2416	87.2	492	2	Q8UTG1 human immun
34	2416	87.2	494	2	Q994R3 human immun
35	2415.5	87.1	491	2	Q8UTR0 human immun
36	2415.5	87.1	491	2	Q9Q3C6 human immun
37	2415	87.1	492	2	Q8UTH0 human immun
38	2414.5	87.1	491	2	Q9QNP8 human immun
39	2413	87.0	488	2	Q8UTM4 human immun
40	2413	87.0	492	2	Q8UTB6 human immun
41	2412	87.0	492	2	Q9WF93 human immun
42	2412	87.0	496	2	Q8AC80 human immun
43	2412	87.0	500	2	Q8AMN8 human immun
44	2411.5	87.0	491	2	Q7SLX9 human immun
45	2410.5	87.0	489	2	Q7SQ55 human immun

## ALIGNMENTS

RESULT 1  
Q9WF75 PRELIMINARY: PRT: 492 AA.  
ID Q9WF75  
AC Q9WF75  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag polyprotein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Nung'u T.P., Rahman M., Makheima M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana";  
RL J. Virol. 73:4427-4432(1999).  
DR EMBL; AF110965; AAD17055.1; -.  
DR PIR; S49086; S49086.  
DR HSSP; Q70622; 1HVN.  
DR GO; GO:0019012; C:violation; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; F:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR010999; Retroviral matrix.  
DR InterPro; IPR000071; Retroviral p17.  
DR InterPro; IPR008916; Retroviral capsid C.  
DR InterPro; IPR008919; Retroviral capsid N.  
DR InterPro; IPR001878; Zm1\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; ze\_CCHC; 2.  
DR SMART; SM00343; ZNF\_C2HC; 2.  
DR PROSITE; PS50158; ZF\_CCHC; 2.









OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OK NCBI\_TaxId=11676;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=2198475; PubMed=11991972;  
DOI=10.1128/JVI.76.11.5435-5451.2002;  
RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,  
Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,  
Foley B.T., Gaoekwe S., Rybak N., Gasetstwe S., Vamberg F.,  
RA Marink R., Lee T.-H., Essex M.,  
RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:  
RT consensus sequence for an AIDS vaccine design?";  
RU J. Virol. 76:5435-5451(2002).  
RN [2]  
RN SEQUENCE FROM N.A.  
RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,  
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gasetstwe S., Vamberg F.,  
RA Marink R., Lee T.-H., Essex M.,  
RU Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF431100; AAL34787.1; -.  
DR HSP; P24736; INCP.  
DR GO; GO:0019012; Cytolysis; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF00098; ZF-CHC; 2.  
DR SMART; SM00343; ZNF\_C2HC; 2.  
DR PROSITE; PS00158; ZF\_C2HC; 2.  
KW AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 492 AA; 55149 MW; 61B3F3B5E87E9A1B CRC64;

## Alignment Scores:

Pred. No.: 3,62e-101 Length: 492  
Score: 2459.00 Matches: 462  
Percent Similarity: 97.36% Conservative: 18  
Best Local Similarity: 93.71% Mismatches: 11  
Query Match: 88.71% Indels: 2  
DB: Gaps: 2

US-09-475-704a-3 (1-1479) x Q8UT62 (1-492)

QY 1 ATGGGCGCCCGCGCAGCATCTGGCGCGCGCAGCTGGAGCCCTGGAGGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleuAlaArgIlyIyLeuAspLysrTrpIuArgIleArg 20  
QY 61 CTGGCGCCCGCGCAGAAAGTGTACATGATGAAGCACTGTGTGGCGCAGCGCAG 120  
DB 21 LeuAlaProGlyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 40  
QY 121 CTGGAAGAGTTGCGCTGAACCCCGCGCTGTGTGAGACCGAGCGCGCTGCAAGCATC 180  
DB 41 LeuGluArgPheAlaLeuAlaProGlyLeuLeuGluThrAlaGlyIyCysIyIyIyIyIy 60  
QY 131 ATGGCGCAGCTCACCGCGCGCTGGAGACCGCGCAGGAGAGCTGAAGAGCTGTTCAC 240  
DB 61 IleArgIleuAlaProAlaLeuGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 80  
QY 241 ACCGTGGCCACCTGTACTGCTGTGACGAGAAAGATGAGTCCGCGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrIleuIyCysValAlaIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 100  
QY 301 CTGCAAGATGAGAGAGAGAGCAACAGTCCAGCAGAGATCCAGAGCGCGCAGGCC 360  
DB 101 LeuAspLysIleGlyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 120  
QY 361 GCCGCAAGGGGAGAGTGAAGCAACTACCCATCGTCAGACACCTGAGGGCGCAGATG 420  
DB 121 AlaAspLys--GlnValSerIleuIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 139  
QY 421 GTGCACAGGCGCATGAGCCCGCGCAGCCTGGAAGCGCTGGGTGAAGGTATCGAGAGAAG 480

DB 140 ValHisGlnProLeuSerProArgThrIleuAlaIyIyIyIyIyIyIyIyIyIyIyIy 159  
QY 481 GCGTTACGCGCGAGGTATCCCATGTTACCGCGCTGACGAGGAGCGCAGCCCCCAG 540  
DB 160 AlaPheSerProGluValIleProMetPheThrAlaIyIyIyIyIyIyIyIyIyIyIyIy 179  
QY 541 GACCTGAACAGATTTGAACACCGTGGCGCGCCAGCGCGCGCCATGCAATGCTGAAG 600  
DB 180 AspLeuAlaThrIleuAlaIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 199  
QY 601 GACACCATCAACAGAGAGCGCGCAGTGGAGACGCGCGCAGCAGCGCGTGCAGCGCGCGCC 660  
DB 200 AspThrIleuAlaGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 219  
QY 661 ATGCGCGCGCGCAGATCGCGAGCGCGCGCGCGCAGCATCGCGCGCAGCAGCAGC 720  
DB 220 ValAlaProGlyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 239  
QY 721 CTGCAAGAGCATGCGCTGATGACGAGCAACCGCGCGCATCCCGTGGGAGCATCTAC 780  
DB 240 LeuGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 259  
QY 781 AAGCGGTGATCATCTGGCGCTGAACAAGATCGTGCAGATGACAGCGCGTGAAGATC 840  
DB 260 LysArgTrpIleValIleuGlyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 279  
QY 841 CTGACATCAACAGAGGCGCGCAGAGAGCGCTTCCGCGACTACGTGAGACCGCTTCAAG 900  
DB 280 LeuAspIleIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 999  
QY 901 ACCCTGGCGCGCGCAGCAGACCGCAGAGGAGTGAAGATGAGTGAAGACCGCATCGTG 960  
DB 300 ThrIleuAlaGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 319  
QY 961 GTGCAGAACGCGCAACCGCGCATGCAAGCATCTGTGCGCGCTTCCGCGCGCGCAGC 1020  
DB 320 ValGlnAlaIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 339  
QY 1021 CTGAGAGATGATGACCGCTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1080  
DB 340 LeuGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 359  
QY 1081 CTGGCGCAGGAGTGAAGCAGAGCGC--AACACGAGCGATGATGACAGAGCAACTTC 1137  
DB 360 LeuAlaGlnAlaIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 379  
QY 1138 AAGGCGCGCGCGCATGCTGAAGTCTTCACTGCGGCAAGAGGCGCAGCGCGC 1197  
DB 380 LysGlyProAlaGlyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 399  
QY 1198 AACTGCGCGCGCGCGCAGAAAGGCTGTGGAAGTGGCGCAGAGGCGCAGCATG 1257  
DB 400 AsnCysArgAlaProArgIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 419  
QY 1258 AAGCATCAGCAGGCGCGCAGCAACTCTGTGGCAAGATCTGGCGCAGCAGAGGCC 1317  
DB 420 LysAspCysThrGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 439  
QY 1318 GCGCGCGCGCACTTCTGCAAGCGCGCGCGCAGCGCGCGCGCGCGCGAGAGCTTC 1377  
DB 440 ArgProGlyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 459  
QY 1378 GCGTTGAGAGACACCGCGCGCGCGCAGAAAGAGAGAGCAAGAGCGCGGAGACCTGAGC 1437  
DB 460 LysPheGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 479  
QY 1438 AGCCTGAAGAGCTGTTGCGCAAGCAGCGCGCTGAGCGCAG 1476  
DB 480 SerIleuIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 492

RESULT 6  
Q8AFG0 PRELIMINARY; PRT; 493 AA.



AC OBAFG0;  
DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_Taxid=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=982A445;  
RX MEDLINE=2267157; PubMed=12680693; DOI=10.1023/A:1022378022104;  
RA Hunt G.M., Papathanasopoulos M.A., Gray G.E., Tiemeisen C.T.;  
RT "Characterisation of near-full length genome sequences of three South  
RT African human immunodeficiency virus type 1 subtype C isolates.";  
RL Virus Genes 26:49-56(2003).  
DR EMBL; AY158533; J0577379.1; -.  
DR HSP; Q07622; 1HVN.  
DR GO; GO:0019012; C:viral; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR000999; Retroviral matrix.  
DR InterPro; IPR000071; Retroviral p17.  
DR InterPro; IPR008916; Retroviral capsid C.  
DR InterPro; IPR008919; Retroviral capsid N.  
DR pfam; PF00540; Gag\_p17; 1.  
DR pfam; PF00607; Gag\_p24; 1.  
DR pfam; PF00098; zf\_CCHC; 2.  
DR SMART; SM00343; ZNF\_C2HC; 2.  
DR PROSITE; PS50158; zf\_CCHC; 2.  
DR AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 493 AA; 55273 MW; DB51A1997BC32EDD CRC64;  
  
Alignment Scores:  
Pred. No.: 5 166-101 Length: 493  
Score: 2455.50 Matches: 459  
Percent Similarity: 97.16% Conservative: 20  
Best Local Similarity: 93.10% Mismatches: 13  
Query Match: 88.58% Indels: 1  
Gaps: 1  
  
US-09-475-704a-3 (1-1479) x OBAFG0 (1-493)  
QY 1 ATGGGCGCGCGCGCGCATCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAserThrTrrpIuylsileArg 20  
QY 61 CTGGCGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACCTGATGGCGCGCGCGAG 120  
DB 21 LeuArgProGlyGlyLeuValysHlyMetIleValIleValIleTrrpIaSerArgIu 40  
QY 121 CTGGAGAGTGGCGCTGAACCCCGCGCTGCGAGACGAGAGCGCTGGAGCGCGAGATC 180  
DB 41 LeuGluArgPheAlaLeuAserProGlyLeuLeuGluThrSerGlnGlyCylsValSglnIle 60  
QY 181 ATCCGCGAGCTGACCGCGCGCGCTGCGAGACGCGCGAGAGAGCTGATTCAC 240  
DB 61 LeuylsGlnIleuHleProAlaLeuGlnThrGlyThrGlnGluLeuArgSerIleuTrpAse 80  
QY 241 ACCGTGGCGACCTGTATCTGCTGTGCAAGAAAGATGAGTCCCGCGACCAAGAGGCG 300  
DB 81 ThrValAlaIleThrLeuTyrCylsValIleSglnArgIleGlnIleArgAserThrIysGlnAla 100  
QY 301 CTGGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 101 LeuAserIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120  
QY 361 GCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

DB 121 AlaAserIleGlyLeuValSerGlnAserIleuThrProIleValIleGlnAserIleGlnIleMet 140  
QY 421 GTGGACACAGCGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
DB 141 ValHleGlnProIleuSerProArgThrLeuAserAlaTrrpValIleValIleGlnGlnIlys 160  
QY 481 GCCTTCAGCGCGCGAGGTATCCCGCATGTTCACCGCGCTGAGAGAGAGAGAGAGAGAGAGAG 540  
DB 161 AlaPheSerProGlnValIleProMetPheThrAlaLeuSerGlnIleValAlaThrProGln 180  
QY 541 GACCTGAACAGATGTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
DB 181 AserLeuAserThrMetLeuAserThrValIleGlyIleSglnAlaAlaMetGlnMetIleuIys 200  
QY 601 GACACCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 201 AserThrIleAserGlnGlnIleAlaIleGlnTrrpAraPheHleIleProValHleIleGlyPro 220  
QY 661 ATCGCGCGCGCGCGAGTGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
DB 221 IleAlaProGlyGlnMetArgGlnProAlaGlySerAserIleAlaGlyThrThrSerAse 240  
QY 721 CTGGAGAGAGAGATCGCTGTGATGACCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
DB 241 LeuGlnGlnGlnIleAlaIleThrMetThrSerAseProIleProValGlyAserIleTyr 260  
QY 781 AACGGGTGATATCTGGGCTGAAACAGATGCGCGAGATGATACAGCGCGCGCGCGCGCGCG 840  
DB 261 IysArgTrrpIleIleLeuGlyLeuAserIleValArgMetIleSerProValSerIle 280  
QY 841 CTGGACATCAAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
DB 281 LeuAserIleIysGlnGlyProIysGlnProPheAraPheIleValAserArgPheIleIys 300  
QY 901 ACCCTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 301 ThrLeuArgAlaGlnGlnIleAlaIleThrGlnIleValIleAserThrMetThrAserThrLeuLeu 320  
QY 961 GTGGAGAACCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020  
DB 321 ValGlnAserAlaAserProAserPylsThrIleLeuArgAlaLeuGlyProIleValAser 340  
QY 1021 CTGGAGAGATGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080  
DB 341 LeuGlnGlnMetThrAlaCylsGlnIleValGlyIleProSerHleIleValAserVal 360  
QY 1081 CTGGCGCGAGCGATGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137  
DB 361 LeuAlaGlnAlaMetSerGlnAlaIleIleAserAseIleMetMetGlnIleAserAsePhe 380  
QY 1138 AAGGCGCGCGCGCGCGCATGTCTCAAGTCTTCACTGCGCGCGAGAGAGAGAGAGAGAGAG 1197  
DB 381 IysGlyProArgArgThrValIleCylsPheAseIleValIleGlnIleIleAlaIlys 400  
QY 1198 AACTGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
DB 401 AseIleCylsAraPheIleProArgIleIleCylsIleValIleCylsIleValIleIleMet 420  
QY 1258 AAGGACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317  
DB 421 IysAserCylsThrGlnIleArgGlnAlaAseIleuGlnIleIleTrrpProSerHleIysGly 440  
QY 1318 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377  
DB 441 ArgProGlyAsePheIleuGlnIleSerArgProGlnProThrAlaProProAlaIleIleSerPhe 460  
QY 1378 CGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
DB 461 ArgPheGlnGlnIleAserThrProValProIysGlnGlnIleIleValIleArgGlnProLeuThr 480  
QY 1438 AGCTGAAGAGCGTGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476  
DB 481 SerIleuIysSerIleuPheGlySerAserProIleuSerGln 493



DB 460 ArgpneugluThrhPrroAlaProlySGIngluProleIleugluArgluProleuthr 479  
QY 1438 ACCCTGAAGAGCCTGTTGCGCAACGACCCCTGAGCCAG 1476  
DB 480 SerleuLySerleupheglSerleuProleuSerGln 492  
RESULT 8  
Q8UT89 PRELIMINARY, PRT, 492 AA.  
AC Q8UT89;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxId=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2198475; PubMed=1191972;  
RX DOI=10.1128/JVI.76.11.5435-5451.2002;  
RA Novitsky V., Smith U.R., Gilbert P., Chigwedere P.,  
RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,  
RA Foley B.T., Gaoletwe S., Rybak N., Gaseitsiwe S., Vamberg F.,  
RA Marlink R., Lee T.-H., Essex M.,  
RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:  
RT consensus sequence for an AIDS vaccine design?";  
RL J. Virol. 76:5435-5451 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,  
RA Chang S.-Y., Peter T., Thior I., Rybak N., Gaseitsiwe S., Vamberg F.,  
RA Marlink R., Lee T.-H., Essex M.,  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF443097; AML34760.1; -  
DR HSSP: P24736; INCP.  
DR GO: GO:0019012; C:virion; IEA.  
DR GO: GO:0003676; F:nucleic acid binding; IEA.  
DR GO: GO:0005198; F:structural molecule activity; IEA.  
DR GO: GO:0016032; P:viral life cycle; IEA.  
DR Pfam: PF00540; Gag\_p17; 1.  
DR Pfam: PF00507; Gag\_p24; 1.  
DR Pfam: PF00098; z1-CHC; 2.  
DR SMART: SM00343; Znf\_C2HC; 2.  
DR PROSITE: PS50158; ZF\_C2HC; 2.  
KW AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 492 AA; 55166 MW; 82973840D175A5D4 CRC64;  
Alignment Scores:  
Pred. No.: 3,4e-100 Length: 492  
Score: 2437.00 Matches: 457  
Percent Similarity: 97.16% Conservative: 22  
Best Local Similarity: 92.70% Mismatches: 12  
Query Match: 87.91% Indels: 2  
DB: 2 Gaps: 2  
US-09-475-704A-3 (1-1479) x Q8UT89 (1-492)  
QY 1 ATGGGGGGGGGGGGCGCATCTCGGGCGGGCGCAAGCTGAGCCCTGGAGGGCGATCCGC 60  
DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLeuAspArgTrpGlnuylleArg 20  
QY 61 CTGGCGCCCGGGCGCAAGAGTCTACATGATGAGACACCTGGTGGCGACGCGCGAG 120  
DB 21 LeuArgProGlyGlyLeuYelYelYelYelYelYelYelValTrpAlaSerArgGlu 40  
QY 121 CTGAGAGATTGGCCCTGAACCCCGGCGCTGGTGAAGACGAGGGCTGACAGCATC 180  
DB 41 LeuGluArgPheAlaLeuAsnProGlyLeuLeuGluThrSerGlnuGlyCysyluSglnIle 60  
QY 181 ATCCGCGACCTGACCCCGCCCTGACAGACCGGACGAGAGAGCTGAAGAGCTGTTCAC 240

DB 61 MetLysGlnLeuGlnProAlaIleIleGlnThrGlyThrGlnGluLeuArgSerleuPheAsn 80  
QY 241 ACCGTGGCCACCTGTACTGCGTGCACAGAAAGATCAGAGTCCGCGACACCAAGAGGCC 300  
DB 81 ThrValAlaThrLeuTrpCysValHisGlnGlyIleAspTrpGlnuyluAla 100  
QY 301 CTGGAAGAATGAGAGAGAGCAAGAGTCCAGAGAGATGACAGAGCCGAGGCC 360  
DB 101 LeuAspLysValGlnGlnGlnuGlnuLysSerGlnuLysMetGlnuAlaGluAla 120  
QY 361 GCCGCAAGGGCGAAGTGAAGCCAGAACTACCCCATGTGCAAGACCTGAGGCGCAGATG 420  
DB 121 AlaAspLys---LysValSerGlnuThrProIleValGlnAsnLeuGlnuGlnuMet 139  
QY 421 GTGCACCAAGGCCATGACCCCGGACCTGAACGCTGGGTGAAGTGTACAGAGAG 480  
DB 140 ValHisGlnuAlaIleSerProArgThrLeuAsnAlaTrpValuValIleGluAspLys 159  
QY 481 GCCTTCAGCCCGGAGTATCCCATGTCACCGGCTGAGGAGGGCGACCCCGCAG 540  
DB 160 AlaPheSerProGluValIleProMetPheThrAlaLeuSerGlnuGlyAlaThrProGln 179  
QY 541 GACCTGAACACGATGTTGAACAACCGTGGGCGGCGCACAGGCCCATGAGATGCTGAAG 600  
DB 180 AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnuAlaIleMetGlnuMetLys 199  
QY 601 GACACCATCAACAGAGAGCGCGCGAGTGGACCGCGTGCACCCGCTGACCGCGGCC 660  
DB 200 AspTrpIleAsnGlnuGluAlaIleAspArgLysLeuAspProValHisIleGlyPro 219  
QY 661 ATCCGCGCCCGGCAATGCGGAGGCCCGGCGGACGATGCGGCGGCGACCGAC 720  
DB 220 ValAlaProGlyGlnuMetArgIleProArgLysSerAspIleAlaGlyThrThrSerTrp 239  
QY 721 CTGCAGAGACGATGCGCTGATGACCAACACCCCATCCCGTGGGCGCACATCTAC 780  
DB 240 LeuGlnGlnuGlnuIleAlaTrpMetThrSerAsnProIleProValGlyAspIleTrp 259  
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGGATPACAGCCCGCTGACATC 840  
DB 260 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIle 279  
QY 841 CTGCACATCAAGAGAGCGCGCAAGAGCCCTTCCGCGCTGAGACCGCTTCTTCAAG 900  
DB 280 LeuAspIleLysGlnuGlyProLysIleProPheArgAspTrpValAspArgPheLys 299  
QY 901 ACCCTGCGCGCGGACAGAGACAGCAGAGAGTGAAGAATGGATGACCGACACCTGCTG 960  
DB 300 ThrLeuAlaGlnuGlnuSerThrGlnuGlnuValuLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGCAGAAAGCCCAACCCCGATGCAAGACATCTGCGCGCTTGGCGCCGCGCGCAC 1020  
DB 320 ValGlnuAlaAsnProAspArgLysThrIleLeuValAlaLeuGlySerIleAlaSer 339  
QY 1021 CTGCAGAGATGATGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
DB 340 LeuGlnuLysMetThrAlaCysGlnuGlyValGlyGlyProSerHisIleValAlaTrpVal 359  
QY 1081 CTGGCGGAGGCGATGAGCCAGGCC---AACACAGAGTGAATGATGACAGAGCAATTC 1137  
DB 360 LeuAlaGluAlaMetSerGlnuAlaAsnAsnValGlnuIleMetMetGlnuLysAsnAsnPhe 379  
QY 1138 AAGGGCGCCCGCGCATGCTCAAGTGTTCATCTGCGGCAAGAGAGCGCATCGCCGCG 1197  
DB 380 LysGlyProArgArgIleIleLysCysPheAsnLysGlyLysGlnuGlyHisIleuAlaArg 399  
QY 1198 AACTGCGCGCGCGCGCGGAAAGAGGCTGCGAGAGTCCGGAAGAGAGCGCAACGATG 1257  
DB 400 AsnCysAspArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnuGlyHisIleMet 419  
QY 1258 AAGGACTGACCGGAGCGCGCAATTCCTGGCGCAAGATCTGGCGGCGGACCAAGGGCG 1317





QY 1138 AAGGCGCCCGGCGCATCGTCAAGTCTTCAACTGCGGAGAGGAGCCCATCGCCCGC 1197  
 |||||  
 DB 380 LybGlyProArgrArgIleVallyscYsphenAcYsGlylybGlyluGlystleAlaArg 399  
 |||||  
 QY 1198 AACTGCGCGCGCCCGCGCAAGAGGCGTGTGGAAGTGTGCGGAGAGGAGCCCATGATG 1257  
 |||||  
 DB 400 AsnCyArgAlaProArgrArglybGlyscYsTrpLyCyGlylybGlyluGlystleAlaMet 419  
 |||||  
 QY 1258 AAGGATGAGCGAGGCGGAGCGCAACTCTCGGAGAGATCTGGCGCCGCGCAAGAGGC 1317  
 |||||  
 DB 420 LybAspCystrHrdLunArglnAlaAsnPhenLeuGlyLysLettPrProSerHlsbYsGly 439  
 |||||  
 QY 1318 CGCCCGCGCAACTCTCGGAGAGCGCCCGGAGCGCCCGCGCGCGCGAGAGCTTC 1377  
 |||||  
 DB 440 ArgProGlyAsnPhenLeuGlnAsnArgProGlnProThrAlaProProAlaGlySerPhe 459  
 |||||  
 QY 1378 CGCTTGAGAGAGACCAACCCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
 |||||  
 DB 460 ArgPheGlyGlnGlyThrThrProAlaProLysGlnGlnProIleGlnLysGlnProLeuThr 479  
 |||||  
 QY 1438 AGCGTGAAGAGCGCTGTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476  
 |||||  
 DB 480 SerLeuLysSerLeuPheGlySerAspProLeuSerGln 492  
 |||||

## RESULT 11

Q8UTH9 PRELIMINARY; PRT; 499 AA.

AC Q8UTH9; 01-MAR-2002 (TRENBLREL. 20, Created)

DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)

DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)

DE Gag protein.

GN Name=gag;

OS Human immunodeficiency virus 1.

OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.

OX NCBI\_Taxid=11676;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21988475; PubMed=1191972;

RA DOI=10.1128/JVI.76.11.5435-5451.2002;

RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chigwedere P.,

RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,

RA Foley B.T., Gaoletke S., Rybak N., Gaseletski S., Vanberg F.,

RA Marink R., Lee T.-H., Essex M.;

RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:

RT consensus sequence for an AIDS vaccine design?";

RL J. Virol. 76:5435-5451 (2002).

RN [2]

RP SEQUENCE FROM N.A.

RA Novitsky V.A., McLane M.F., Chigwedere P., Ndung'u T., Klein I.,

RA Chang S.-Y., Peter T., Thior I., Rybak N., Gaseletski S., Vanberg F.,

RA Marink R., Lee T.-H., Essex M.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL AF440871; AAL34670.1; -.

DR HSSP; Q70622; IHVN.

DR GO; GO:0019012; C:virion; IEA.

DR GO; GO:0003676; P:nucleic acid binding; IEA.

DR GO; GO:0005198; P:structural molecule activity; IEA.

DR GO; GO:0016032; P:viral life cycle; IEA.

DR Pfam; PF00540; Gag\_p17; 1.

DR Pfam; PF00607; Gag\_p24; 1.

DR Pfam; PF00098; zF-CHC; 2.

DR PRINTS; PR00939; C2HCZNFINGER.

DR PRINTS; PR00234; HIVMATRIX.

DR SMART; SM00343; ZNF\_C2HC; 2.

DR PROSITE; PS50158; ZF-CHC; 2.

KW AIDS; Core protein; Metal-binding; Polypeptide; Zinc; Zinc-finger.

SO SEQUENCE 499 AA; 55793 MW; E573AF16EEBFB7F CRC64;

## Alignment Scores:

Pred. No.: 5,37e-100 Length: 499  
 Score: 2432.50 Matches: 455  
 Percent Similarity: 95.99% Conservative: 24

Best Local Similarity: 91.18% Mismatches: 13  
 Query Match: 87.75% Indels: 7  
 DB: 2 Gaps: 2

US-09-475-704A-3 (1-1479) x Q8UTH9 (1-499)

QY 1 ATGGGCGCCCGGCGGAGATCTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
 |||||  
 DB 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspThrTrpGlnLysIleArg 20  
 |||||  
 QY 61 CTGGCGCCCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
 |||||  
 DB 21 LeuArgProGlyGlyLysLysIleArgMetIleLysIleValThrAlaSerArgln 40  
 |||||  
 QY 121 CTGGAGAGAGTGGCCCTGAACCCCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 180  
 |||||  
 DB 41 LeuGlnArgPheAlaLeuAsnProLysLeuLeuGlnThrSerGlnGlyCysLysGlnIle 60  
 |||||  
 QY 181 ATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 |||||  
 DB 61 MetLysGlnLeuGlnProAlaLeuGlnThrGlyThrGlnLysLeuArgSerLeuThrAsn 80  
 |||||  
 QY 241 ACCGTGACCGACCTGTACTGTGCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
 |||||  
 DB 81 ThrIleAlaThrLeuThrCysValHisGlnSerIleGlnValArgAspThrIleGlnAla 100  
 |||||  
 QY 301 CTGGAGAGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 |||||  
 DB 101 LeuAspLysIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120  
 |||||  
 QY 361 GCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
 |||||  
 DB 121 AlaAspLysGlyLysValSerGlnAsnThrProIleValGlnAsnLeuGlnGlnMet 140  
 |||||  
 QY 421 GTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
 |||||  
 DB 141 ValHisGlnAlaIleSerProArgThrLeuAsnAlaThrValLysValIleGlnLys 160  
 |||||  
 QY 481 GCGTTCAGCCCGGAGAGATCCCATGTTCACCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 |||||  
 DB 161 AlaPheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGln 180  
 |||||  
 QY 541 GACCTGAACAGAGATGTGAACCGGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 |||||  
 DB 181 AspLeuAsnThrMetLeuAsnThrValGlyGlnAlaIleValMetLeuLys 200  
 |||||  
 QY 601 GACACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
 |||||  
 DB 201 AspThrIleAsnGlnGlnAlaIleGlnLysProArgThrAspThrIleProValHisAlaGlyPro 220  
 |||||  
 QY 661 ATCGCGCCCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 |||||  
 DB 221 IleAlaProGlyGlnMetArgGlnProArgLysSerPheIleAlaGlyThrThrSerThr 240  
 |||||  
 QY 721 CTGGAGAGAGAGATCGCTGTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 |||||  
 DB 241 LeuGlnGlnGlnIleAlaThrMetThrSerAsnProProValProValGlyGlnIleThr 260  
 |||||  
 QY 781 AAGCGGTGATCATCTGTGGCGCTGAACAAGATCGTGGAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGATC 840  
 |||||  
 DB 261 LysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetLysSerProValSerIle 280  
 |||||  
 QY 841 CTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 |||||  
 DB 281 LeuAspIleLysGlnGlyProLysGlnProArgThrAspThrValAlaAspArgPhePheLys 300  
 |||||  
 QY 901 ACCGTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 |||||  
 DB 301 ThrLeuArgAlaGlnGlnAlaSerGlnAspValLysAsnThrMetThrAspThrLeuLeu 320  
 |||||  
 QY 961 GTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 |||||  
 DB 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlnProGlyAlaSer 340  
 |||||





QY 901 ACCCTGCGCGCGAGAGAGCAACCCGAGAGTGAAGAACTGATGACCCGACCTCTGCTG 960  
 Db 300 ThrleuAAGAlaGlnGlnAlaThrglnaPValVAsnTrpMetThrAphThleu 319  
 QY 961 GTGCAGAAAGCCCAACCCGATGCAAGACCATCTCTGCGCTCTGCGCCCGCGCGCAGC 1020  
 Db 320 ValGlnaAAlaAsnProAspCysIsthrIleuAAGAlaLeuGlyProGlyAlaThr 339  
 QY 1021 CTGAGAGAGATATGACCGCTCTGCGAGCGCGTGGCGCGCCGACCAAGCGCCGCTG 1080  
 Db 340 LeuGlnGlnMetMetThrAlaCysGlnGlnValGlyIleProSerIsthrAlaArgVal 359  
 QY 1081 CTGCGCGAGCGATGAGCGAGCGC---AACACCGAGTATATGACAGAGCAACTTC 1137  
 Db 360 LeuAlaGlnAlaMetSerGlnAlaAsnAlaAsnIleMetMetGlnArgSerAsn 379  
 QY 1138 AAGGCGCGCGCGCATCTGCAAGTCTTCAACTGCGCGAGAGAGGCGCACATCGCGCGC 1197  
 Db 380 LysGlyProArgLysIleIleLysCysPheAsnGlyLysGlnGlyIsthrAlaArg 399  
 QY 1198 AACTGCGCGCGCGCGCGAGAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCAACGATG 1257  
 Db 400 AasnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyIsthr 419  
 QY 1258 AAGGACTGACCGAGCGCGCGCAACTCTCTGCGCGAGATCTGCGCGCGCAAGGCGC 1317  
 Db 420 LysAspCysThrGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerIsthr 439  
 QY 1318 CGCGCGCGCACTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCTTC 1377  
 Db 440 ArgProGlnAlaAsnPheLeuGlnaAsnArgProGlnProThraProProAlaGln 459  
 QY 1378 CGCTTGAGAGAGCAACCCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1437  
 Db 460 ArgPheGlnGlnThrThrProAlaProLysGlnGlnGlnLysAspArgLysProLeuThr 479  
 QY 1438 AGCCTGAGAGAGCGCTGTGGCAAGAGAGCGCGCGCTGAGAGCAG 1476  
 Db 480 SerLeuLysSerLeuPheIsthrAspProLeuSerGln 492  
 RESULT 13  
 Q8UT53 PRELIMINARY; PRT; 492 AA.  
 AC Q8UT53:  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gag protein.  
 GN Name=gag;  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11676;  
 RX DOI=10.1128/JVI.76.11.5435-5451.2002;  
 RX MEDLINE=21988475; PubMed=11991972;  
 RA Novitsky V., Smith U.R., Gilbert P., McLane M.F., Chiswedere P.,  
 RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,  
 RA Foley B.T., Golekwe S., Rybak N., Gasettsiwe S., Vanberg F.,  
 RA Marink R., Lee T.-H., Essex M.;  
 RT "Human immunodeficiency virus type 1 subtype C molecular phylogeny:  
 RT consensus sequence for an AIDS vaccine design?";  
 RL J. Virol. 76:5435-5451(2002).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP Novitsky V.A., McLane M.F., Chiswedere P., Ndung'u T., Klein I.,  
 RA Chang S.-Y., Peter T., Thior I., Rybak N., Gasettsiwe S., Vanberg F.,  
 RA Marink R., Lee T.-H., Essex M.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF443101; AAL34796.1; -.  
 DR HSSP, Q70622; 1HVN.  
 DR GO, GO:001912; C:virion; IEA.  
 DR GO, GO:0003676; F:nucleic acid binding; IEA.

DR GO, GO:0005198; F:structural molecule activity; IEA.  
 DR GO, GO:0016032; F:viral life cycle; IEA.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF00098; zF-CHC; 2.  
 DR SMART; SM00343; zF\_C2HC; 2.  
 DR PROSITE; PS0058; zF\_CCHC; 2.  
 KW AIDS; Core protein; Polyprotein.  
 SQ SEQUENCE 492 AA; 55087 MW; 9880F1A142680C9 CRC64;  
 Alignment Scores:  
 Pred. No.: 7,68e-100 Length: 492  
 Score: 2429.00 Matches: 457  
 Percent Similarity: 96.75% Conservative: 20  
 Best Local Similarity: 92.70% Mismatches: 14  
 Query Match: 87.63% Indels: 2  
 DB: 2 Gaps: 2  
 US-09-475-704a-3 (1-1479) x Q8UT53 (1-492)  
 QY 1 ATGGCGCGCGCGCGAGCATCTGCGCGCGCGCAAGCTGGAAGCGCTGGAAGCGCATCCGC 60  
 Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyGlyLysLeuAspTrpTrpGlnLysIleArg 20  
 QY 61 CTGCGCGCGCGCGCGCAAGTGTACTATGATGAAGCACTGTGTGGCGCGCGCGAG 120  
 Db 21 LeuArgProGlyGlyLysLysArgGlyMetLeuLysShsLeuIleTrpThrSerArgGln 40  
 QY 121 CTGGAAGATTCGCGCTGACCCCGCGCGCTGCTGAGAGACAGCGAGGCGCTGACAGATC 180  
 Db 41 LeuGlnArgPheAlaLeuAsnProAspLeuLeuGlnTrpSerGlnGlyCysLysGlnIle 60  
 QY 181 ATCGCGCGCTGACCGCGCGCTGACAGCGCGAGAGAGCTGGAAGCGCTGTTCAAC 240  
 Db 61 IleGlnGlnLeuGlnProAlaLeuGlnThrGlyThrGlnGlnLeuArgSerLeuTrpAsn 80  
 QY 241 ACCGCGCGCGCGCTGATCTGCGTGAAGAGATTCGAGCGCGAGAGAGAGAGAGAGAGC 300  
 Db 81 ThrValAlaTrpLeuTrpCysValHisLysGlyIleGluValGlnAspTrpLysGlnAla 100  
 QY 301 CTGGAAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 360  
 Db 101 LeuAspLysIleGlnGlnGlnGlnAsnLysSerGlnGlnLysTrpGlnGlnAlaGlnAla 120  
 QY 361 GCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
 Db 121 Ala--AlaGlyLysValSerGlnAsnTrpProIleValGlnAsnLeuGlnGlyGlnMet 139  
 QY 421 GTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
 Db 140 ValHisGlnSerIleSerProArgTrpLeuAsnAlaTrpValLysValIleGlnGlnLys 159  
 QY 481 GCCTTCAGCGCGAGAGTATCCCATGTTCAACCGCGCTGAGAGAGAGAGAGAGAGAGAG 540  
 Db 160 AlaPheSerProGlnValIleProMetPheThrAlaLeuSerGlnGlyAlaThrProGln 179  
 QY 541 GACTGAACAGATGTTGAACAACGCTGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 Db 180 AspLeuAsnTrpMetLeuAsnThrValGlyGlnGlnAlaAlaMetGlnMetLeuLys 199  
 QY 601 GACACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 660  
 Db 200 AspTrpIleAsnGlnGlnAlaAlaGlnTrpAspArgLeuHisProValHisAlaGlyPro 219  
 QY 661 ATCGCGCGCGCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 Db 220 IleAlaProGlnGlnLeuArgGlnProArgLysIleAspIleAlaGlyThrThrSerThr 239  
 QY 721 CTGCAAGAGAGAGATCGCTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 Db 240 LeuGlnGlnGlnIleAlaTrpMetThrAsnAsnProValProValGlyLysAspIleTrp 259  
 QY 781 AAGGCGTGAATCCTGCGCGCTGAAAGAGATCGGCGAGATGACAGCGCGTGAAGATC 840



```

DB      260  LysArgTrpIleIleLeuGlyLeuAmuYsIleValaArgMetCysSerProValSerIle 279
QY      841  CTGACATCAAGACGAGGCGCCCAAGAGCCCTTCCGACTACGTGACCGCTTCTTCAAG 900
DB      280  LeuAapIleIleGlnGlnIlyProIlysgIuProPheArgAspTyrValaAspArgPhePheIly 299
QY      901  ACCCTGCGCGCGGACGACGACCCGAGAGGTGAAGAATCTGATGACCAACCTGCTG 960
DB      300  ThrLeuArgAlaGlnIleSerSerGlnIlyValaAsnTrpMetThrAspThrLeu 319
QY      961  GTGCAAGAACCAACCCGACTGCAAGACATCTGCGGCGCTGCGCCCGCGCGCCG 1020
DB      320  ValGlnAsnAlaAsnProAspCybIlysnTrIleLeuArgAlaLeuGlyProGlyAlaTr 339
QY      1021  CTGAGAGAGATGATGACCGCTGCGACGAGCGTGGCGCGCCGACCAAGCGCGCGT 1080
DB      340  LeuGlnGlnMetMetThrAlaCysGlnGlnIlyValGlyIlyProSerHisIlysaIaArgAla 359
QY      1081  CTGCGCGGAGCGCATGACCGCAGCC--AACACCGAGCTGATGATGACGAAGCAACTTC 1137
DB      360  LeuAlaGlnAlaMetSerGlnAlaAsnAsnTrpAsnIleMetGlnIlySerAsnPhe 379
QY      1138  AAGGCGCGCGCGGACATGCTCAAGTCTTCAACTGCGGAGAGGCGCACATCGCGCGC 1197
DB      380  LysGlnIlyProArgArgIleValIlyscYsPheAsnCybIlysgIuGlnIlyHisIleAlaArg 399
QY      1198  AACTCGCGCGCGCGCGCGCAAGAGGCTCTGGAAGTGGCGGACGAGGCGCACAGATG 1257
DB      400  AasnCybArgAlaProArgIlysgIyCybTrpIlyscYbGlyIlysgIuGlnIlysgIuMet 419
QY      1258  AAGACTGACACGAGCGGCGCGCAACTTCTGGGCGAAGATCTGGCGCCGACCAAGGCG 1317
DB      420  LysAspCybThrGlnArgIlnAlaAsnPheLeuGlnIlyIleTrpProSerHisIlysgIy 439
QY      1318  CGCGCGCGGCACTTCTGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGACTTC 1377
DB      440  ArgProGlyAsnAsnPheLeuGlnIleSerArgProGlnProThrAlaProProAlaGlnSerPhe 459
QY      1378  CGCTTCGAGAGACCAACCCCGCGCGCGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
DB      460  ArgPheGlnGlnIlyThrThrProAlaProIlysgIuProIlyAspArgIuProLeuThr 479
QY      1438  AGCTTGAGAGCGCTGTTGGCAACGACCCCTGAGCGCAG 1476
DB      480  SerLeuIlySerLeuPheGlySerAspProLeuSerGln 492
RESULT 14
Q7SLX7 PRELIMINARY; PRT; 491 AA.
AC Q7SLX7;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Gag Protein.
GN Name=gag;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_Taxid=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Gupta R.M., Prasad V., Singh B., Seth P.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF531311; AAP87971.1; -.
DR HSSP; 070622; 1HVN.
DR GO; GO:0019012; C:action; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR010999; Retroviral matrix.
DR InterPro; IPR000071; Retroviral p17.
DR InterPro; IPR008916; Retroviral capsid_C.

```

```

DR      InterPro; IPR008919; Retroviral capsid_N.
DR      InterPro; IPR001878; Znf_CCHC.
DR      Pfam; PF00540; Gag_p17; 1.
DR      Pfam; PF00607; Gag_p24; 1.
DR      Pfam; PF00098; zF-CCHC; 2.
DR      SMART; SM00343; Znf_C2HC; 2.
DR      PROSITE; PS50158; zF_CCHC; 2.
KW      AIDS; Core protein; Polyprotein.
SQ      SEQUENCE 491 AA; 54815 MW; 36CD9A56CAEAD7C CRC64;

Alignment Scores:
Pred. No.:      8,94e-100      Length:      491
Score:      2427.50      Matches:      455
Percent Similarity:      96.75%      Conservative:      21
Best Local Similarity:      92.48%      Mismatches:      15
Query Match:      87.57%      Indels:      1
DB:      2      Gaps:      1

US-09-475-704A-3 (1-1479) x Q7SLX7 (1-491)
QY      1  ATGGGCGCGCGCGCGACGATCTGCGCGGCGCAAGCTGAGACGCTGGAGCGCATCCG 60
DB      1  MetGlyAlaArgAlaSerIleLeuArgGlyGlyIlyLeuAspIlystrpGlnIlyAsnArg 20
QY      61  CTGCGCGCGCGCGGCAAGAGTCTACATGATGAGACACCTGCTGTGGCGCGCGCGAG 120
DB      21  LeuArgProGlyGlyIlyLysIlysnIlyrMetLeuIlysnIlyLeuValTrpAlaSerArgIly 40
QY      121  CTGAGAGATTGCGCGCTGAACCCCGCGCTGTGAGAGACCGAGAGGCGCTGCAAGAGATC 180
DB      41  LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnIlyThrGlnIlyLeuArgSerLeuIlyraen 60
QY      181  ATCCGCGAGCTGACCCCGCGCTGCAAGACCGGACGAGAGAGCTGAGAGCTTCAAC 240
DB      61  LeuIlysgIleuGlnProAlaLeuGlnIlyThrGlnIlyLeuArgSerLeuIlyraen 80
QY      241  ACCGTGGCCACCTGTACTGCTGTGACGAGAGAGATCGAGTCCGGGACCAAGAGAGCGC 300
DB      81  ThrValAlaThrLeuIlyrcYsValHisAlaGlyIleGlnValaArgAspThrIlysgIuAla 100
QY      301  CTGACAGATGAGAGAGAGAGAGCAAGAGTCCAGAGAGATCCAGAGGCGCGAGCGC 360
DB      101  LeuAspIlyIleGlnIlyGlnIlysnIlysgIyGlnIlyThrGlnIlyAlaIlysgIy 120
QY      361  GCGGCAAGGCGAGGTGAGCCGACCACTACCCCATGTGACAGAACCTGACGAGGCGCAGATG 420
DB      121  AlaAsp---GlyIlyValSerGlnIlysnIlyrProIleValGlnAsnLeuGlnIlyGlnMet 139
QY      421  GTGCAACGAGCGCATGAGCCCGCGCGACCTGGAACGCGCTGGGTGAAGTATCGAGAGAG 480
DB      140  ValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValIlyValIleGlnIlyIly 159
QY      481  GCGTTCAGCGCGGAGGTGATCCCAATGTTACCGCGCTGAGAGGAGGCGCGACCCCGCAG 540
DB      160  AlaPheSerProGlnIlyAlaIleProMetPheThrAlaLeuSerGlnIlyAlaThrProGln 179
QY      541  GACTGGAACGAGATGTTGAACCGGTGGCGGCGCACAGGCGCGCATGACGATGCTGAAG 600
DB      180  AspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuIly 199
QY      601  GACACCATCAACAGAGAGGCGCGCGAGTGGGACCGCGCTGCAACCCGCTGACCGCGCGCC 660
DB      200  AspTrpIleAsnGlnIlyAlaIleIlyTrpAspArgLeuHisProValHisAlaGlyPro 219
QY      661  ATCGCGCGCGCGGACATGCGGAGCGCGCGCGCGCGACGATCGCGCGCGCGCACGACCC 720
DB      220  IleAlaProGlyGlnMetArgIlyProArgIlySerAspIleAlaGlyThrThrSerThr 239
QY      721  CTGCAAGAGAGATGCGCTGATGACGAGCAACCCCGCATCCCGTGGGCGCAATTCAC 780
DB      240  LeuGlnIlyGlnIlyAlaIlyrMetThrAsnAsnProProValProValGlyAspIlyr 259
QY      781  AAGCGGTGATCATCTCGGCGCTGGAACAAAGATCGCGGAGTATACAGCCCGCTGAGCATC 840

```

Db 260 LysAlgtPDIlelleuLgLyLeuAmnLysIleValAlaMetLysSerProValSerIle 279  
QY 841 CTGGAATCAACAGAGGCGCCCAAGAGACCTTCCGCACTACGTGAGCCGCTTTCTTCAAG 900  
Db 280 LeuAapIleLysGlnGlyProLysGlnProPheArgAspTyrValAspArgPhePheLys 299  
QY 901 ACCCTGGCGGCGGAGAGAGCAACCGAGAGGTGAAGAATTGATGACCGACACCTGCTG 960  
Db 300 ThrLeuArgAlaGlnGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 319  
QY 961 GTGCAAAAGCCCAACCCCGCAAGCAACATCTGCGGCTCTGCGCCCGCGCGCAAG 1020  
Db 320 ValGlnAsnAlaAsnProAspCysLysTrpIleLeuArgAlaLeuGlyProGlyAlaThr 339  
QY 1021 CTGAGAGAGATATGACCGCTGCGCAAGGCGGTGGCGGCGCCCAACCAAGCGCGCTG 1080  
Db 340 LeuGlnGlnMetMetThrAlaCysGlnGlyValGlyGlyProSerThrIleLysAlaArgVal 359  
QY 1081 CTGGCGGAGGAGATGAGCGGCGCAACACGAGGTGATGACGAGAGCAACTTCAAG 1140  
Db 360 LeuAlaGlnAlaMetSerGlnThrGlySerThrIleMetMetGlnArgSerAsnPheLys 379  
QY 1141 GCGCCCGCGCGCATCGTCAAGTGTCTTCACTGCGGCAAGAGGCGCACTCCCGCAAC 1200  
Db 380 GlySerLysArgIleValLysCysPheAsnGlyLysGlyGlnGlyHisIleAlaArgAsn 399  
QY 1201 TGCCCGCGCCCGCAAGAGGCGTGTGAAGTGTGGCGGCAAGAGGCGCAACGATGAAG 1260  
Db 400 CysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGlnGlyHisGlnMetLys 419  
QY 1261 GACTGACAGGAGCGGCGGCAACTTCCGCGGCAAGATGTGGCCGCAACGAGGCGCG 1320  
Db 420 AspCysTrpGlnArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 439  
QY 1321 CCGCGCAACTTCTGTGAGAGCGCGCCCGAGCCCAACCGCCCGCGCGAGACTTCCG 1380  
Db 440 ProGlnAsnPheLeuGlnSerArgProGlnProThrAlaProProAlaGlnSerPheArg 459  
QY 1381 TTGAGAGAGACCAACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Db 460 PheGlnGlnThrThrProAlaProLysGlnGlnLeuLysAspArgGlnProLeuThrSer 479  
QY 1441 CTGAGAGCGCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476  
Db 480 LeuLysSerLeuPheLysSerAspProLeuSerGln 491  
RESULT 15  
Q7SLX3 PRELIMINARY; PRT; 492 AA.  
AC Q7SLX3;  
DT 01-OCT-2003 (TREMBLrel. 25, Created)  
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviruses; Retroviridae; Lentivirus.  
OC NCBI\_TaxID=11676;  
RN NCBI\_TaxID=11676;  
RP SEQUENCE FROM N.A.  
RA Gupta R.M., Prasad V., Singh B., Seth P.;  
RT Submitted (Jul-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AF53135; AAP87975.1; -.  
DR HSP; Q70622; HIV.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPRO00721; Gag\_P24.  
DR InterPro; IPRO10999; Retroviral\_matrix.  
DR InterPro; IPRO00071; Retroviral\_p17.  
DR InterPro; IPRO08916; Retroviral\_capsid\_C.

DR InterPro; IPRO08919; Retroviral\_capsid\_N.  
DR InterPro; IPRO01878; ZnF\_CCHC.  
DR Pfam; PF00540; Gag\_P17; I.  
DR Pfam; PF00607; Gag\_P24; I.  
DR Pfam; PF00098; zF-CCHC; 2.  
DR SMART; SM00343; ZnF\_C2HC; 2.  
DR PROSITE; PS50158; zF\_CCHC; 2.  
KW AIDS; Core protein; Polyprotein.  
SQ SEQUENCE 492 AA; 55105 MM; 78697BC7E49D3DC3 CRC64;  
Alignment Scores:  
Pred. No.: 1,286-99 Length: 492  
Score: 2424.00 Matches: 454  
Percent Similarity: 95.93% Conservative: 18  
Best Local Similarity: 92.28% Mismatches: 20  
Query Match: 87.45% Indels: 0  
DB: 2 Gaps: 0  
US-09-475-704A-3 (1-1479) x Q7SLX3 (1-492)  
QY 1 ATGGGCGCGCGCGCGAGCATCTGCGGCGGCGCAAGCTGACGCTGGAGCGCATCCGC 60  
Db 1 MetGlyAlaArgAlaSerIleLeuArgGlyLysLeuAspLysTrpGlnLysIleArg 20  
QY 61 CTGCGCGCGCGGCAAGAGTGTACATGATGAAGCACTGTGTGGCGAGCGCGAG 120  
Db 21 LeuArgProGlyGlyLysLysArgTrpMetLeuLysHisLeuValTrpAlaSerArgGln 40  
QY 121 CTGAGAAAGTTGCGCTGAACCCCGCGCTGTGAGAGACGAGCGAGGCTGCAAGCATC 180  
Db 41 LeuGlnArgPheAlaLeuAsnProGlyLeuLeuGlnThrSerThrGlyCysLeuGlnIle 60  
QY 181 ATCCGCGAGCTGCAACCCCGCGCTGCAAGCGGAGGAGGAGCTGAAGCTGTTCAC 240  
Db 61 IleLysGlnLeuGlnProAlaLeuGlnThrGlnTrpGlnGlnLeuArgSerLeuPheAsn 80  
QY 241 ACCGTGCGCACCTGTACTGTGCTGCAAGAGATCGAGTCCGCGCAACCAAGAGAGCC 300  
Db 81 ThrValAlaThrLeuLysCysValHisGlnGlyIleGlnValArgAspThrLysGlnAla 100  
QY 301 CTGCAAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 101 LeuAspLysIleGlnGlnGlnGlnAsnLysIleGlnGlnLysThrGlnGlnLysGln 120  
QY 361 GCGGCAAGGCGAGGTGAGAGCAACTACCCCATCTGCGAGAGACCTGAGGCGGAGATG 420  
Db 121 AlaAspGlyAsnGlnValSerIleAsnTrpProIleValGlnAsnLeuGlnGlyIleMet 140  
QY 421 GTGCAACGAGCATCAGCGCCCGCAACCTGAAAGCTGGGTGAAGTATCGAGAGAGAAG 480  
Db 141 ValHisGlnProIleSerProArgThrLeuAsnAlaTrpValLysValIleGlnGlnLys 160  
QY 481 GCTTCAGCGCGGAGGTATCCCATGTTACCGCGCTGAGCGAGGAGGCGCAACCCCGAG 540  
Db 161 AlaPheSerProGlnValIlePheMetPheThrAlaLeuSerGlnGlyAlaThrProGln 180  
QY 541 GACCTGAACAGATGTGAACACGCTGGCGGCGCAACGAGCGCGCATGACAGTGTGAAG 600  
Db 181 AspLeuAsnTrpMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLys 200  
QY 601 GACACCATCAACAGAGAGGCGCGAGTGGAGACGCGTGAACCCCGTCAAGCGCGGCGCC 660  
Db 201 AspThrIleAsnGlnGlnAlaAlaGlnTrpAspArgLeuHisProValHisAlaGlyPro 220  
QY 661 ATGCGCGCGCGAGATCGCGAGCGCGCGCGGCGAGCAATCGCGGAGCAACCAAGCAC 720  
Db 221 IleAlaProGlyGlnMetArgLysProArgGlySerAspIleAlaGlyThrThrSerThr 240  
QY 721 CTGAGAGAGAGATGCGCTGATGACAGCAACCCCGCATCCCGTGGGAGCATCTAC 780  
Db 241 LeuGlnGlnGlnIleAlaTrpMetThrSerAsnProProValProValGlyAspIleLys 260  
QY 781 AAGCGGTGATCATCTGCGGCTGAAACAAGATCTGCGGAGTACAGCCCGCTGAGCATC 840

```
Db 261 LysArgTrpIleIleuGlyLeuAenLysIleValArgMetCysSerProAlaSerIle 280
QY 841 CTGGACATCAAGAGGGCCCAAGAGCCCTTCGGACTACGTGACCCGCTTCTTCAG 900
Db 281 LeuAspIleLeuGlnGlyProLysGluProPheArgAspTyrValAspArgPhePheLys 300
QY 901 ACCCTGCGCGCCGAGAGAGACCCGAGAGGTGAAGACTGGATGACCGACACCCCTGCTG 960
Db 301 ThrLeuArgAlaGluGlnAlaThrGlnAspValLysAsnTrpMetThrAspThrLeuLeu 320
QY 961 GTGCAAAACGCCAACCCCGACTGCAAGACATCTCGCGCTTCGCGCCCGCGCGCACG 1020
Db 321 ValGlnAsnAlaAsnProAspCysLysThrIleLeuArgAlaLeuGlyProGlyAlaSer 340
QY 1021 CTGGAGAGATGATGACCCCTCGCGAGGCGCTGGGGCGCCCAAGCCACGCGCTG 1080
Db 341 LeuGluGluMetMetThrIleCysGlnGlyValGlyGlyProSerHisLysAlaArgVal 360
QY 1081 CTGGCGGAGCGATGAGCGAGCGCAACACGAGGTGATGATGACAGAAAGCAACTTCAG 1140
Db 361 LeuAlaGluAlaMetSerGlnThrGlySerThrIleMetMetGlnArgSerAsnPheLys 380
QY 1141 GGCCTCCGCGCGCATCGTCAAGTGTCTTCACTGCGGCAAGAGGGCCACATCGCCGCAAC 1200
Db 381 GlySerLysArgIleValLysCysPheAsnGlyGlyLysGluGlyHisIleAlaArgAsn 400
QY 1201 TGCCGCGCGCCCGCAAGAGGGCTGCTGAAGTGTGCGGCAAGAGGGCCACCATGTAAG 1260
Db 401 CysArgAlaProArgLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLys 420
QY 1261 GACTGACCGAGCGCGCAAGCTTCCTGGGGCAAGATCTGGCCGCAAGAGGGCGCGC 1320
Db 421 AspCysTrpGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLysGlyArg 440
QY 1321 CCGGCGCAACTCTCTGAGAGCGCGCCCGAGCCGACCGCCCGCCCGCGAGAGCTTCGCG 1380
Db 441 ProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPheArg 460
QY 1381 TTCGAGAGAGACCAACCCCGCGCAAGAGCAAGAGAGAGACCGCGAGACCTTGACCAAGC 1440
Db 461 PheGluGluThrThrProAlaProLysGlnGluLeuLysAspArgGluProLeuThrSer 480
QY 1441 CTGAAGAGCGCTGTGGGCAAGAGACCCCTGAGCCAG 1476
Db 481 LeuLysSerLeuPheIleLysSerAspProLeuSerGln 492
```

Search completed: March 11, 2005, 15:39:19  
Job time : 285.42 secs

**THIS PAGE BLANK (USPTO)**



181 ATCCGCACTGACACCCGCTCTGCAACCCGCAAGAGAGAGTGAAGACCTGTTCAAC 240  
QY 241 ACCGTGGCCACCTGTACTGCGTGCAAGAAAGATGAGTCCGCAACCAAGAGGCC 300  
DB 241 ACCGTGGCCACCTGTACTGCGTGCAAGAAAGATGAGTCCGCAACCAAGAGGCC 300  
QY 301 CTGGAAGAAGTGAAGAGAGAGAAACAAGTCCAGCAGAAATATCAGAGGCCAGGCC 360  
DB 301 CTGGAAGAAGTGAAGAGAGAGAAACAAGTCCAGCAGAAATATCAGAGGCCAGGCC 360  
QY 361 GCCGCAAGAGGAGAGTGAAGCACTAATCCCATCTGTCAGAACTTGAGAGGCCAGATG 420  
DB 361 GCCGCAAGAGGAGAGTGAAGCACTAATCCCATCTGTCAGAACTTGAGAGGCCAGATG 420  
QY 421 GTGACACAGGCCATCAGCCCCCGCACTCTGAACGCTTGAGAGAGGCCAGGCCAG 480  
DB 421 GTGACACAGGCCATCAGCCCCCGCACTCTGAACGCTTGAGAGAGGCCAGGCCAG 480  
QY 481 GCCTTCAGGCCGAGAGTGAATCCCATGTTCAACGCGCTGAGAGAGGCCAGGCCAG 540  
DB 481 GCCTTCAGGCCGAGAGTGAATCCCATGTTCAACGCGCTGAGAGAGGCCAGGCCAG 540  
QY 541 GACCTGAACACGATGTTGAACAACCGTGGCGGCAACAGGCCGCAATGAGTGAAG 600  
DB 541 GACCTGAACACGATGTTGAACAACCGTGGCGGCAACAGGCCGCAATGAGTGAAG 600  
QY 601 GACACCATCAACAGAGAGGCCGCGCAAGTGGAGCCGCTGACCCGCTGACCCGCGCC 660  
DB 601 GACACCATCAACAGAGAGGCCGCGCAAGTGGAGCCGCTGACCCGCTGACCCGCGCC 660  
QY 661 ATGCGCCCGGAGAGTGGAGAGCCGCGCAAGATGAGTGGAGAGGCCAGGCCAG 720  
DB 661 ATGCGCCCGGAGAGTGGAGAGCCGCGCAAGATGAGTGGAGAGGCCAGGCCAG 720  
QY 721 CTGCAAGAGACAGATCGCTGATGAACAACACCCCTCCATCCCGTGGCGCACTTAC 780  
DB 721 CTGCAAGAGACAGATCGCTGATGAACAACACCCCTCCATCCCGTGGCGCACTTAC 780  
QY 781 AAGCGGTGATATCTCTGGGCTGAAACAAGATGAGTGGAGAGGCCAGGCCAG 840  
DB 781 AAGCGGTGATATCTCTGGGCTGAAACAAGATGAGTGGAGAGGCCAGGCCAG 840  
QY 841 CTGGAATCAACAGAGAGGCCGCGCAAGAGAGCTGAGTGAACCGCTTCTTCAAG 900  
DB 841 CTGGAATCAACAGAGAGGCCGCGCAAGAGAGCTGAGTGAACCGCTTCTTCAAG 900  
QY 901 ACCCTGGCGCGAGCAAGAGCAACCAAGAGTGAAGAACTGATGACCGACCTGCTG 960  
DB 901 ACCCTGGCGCGAGCAAGAGCAACCAAGAGTGAAGAACTGATGACCGACCTGCTG 960  
QY 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
DB 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
QY 1021 CTGGAAGAGATGATACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
DB 1021 CTGGAAGAGATGATACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1081 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
DB 1081 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1141 GGGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200  
DB 1141 GGGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200  
QY 1201 TGCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
DB 1201 TGCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1261 GACTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1261 GACTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320

QY 1321 CCCGCAACTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
DB 1321 CCCGCAACTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
QY 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
QY 1441 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
DB 1441 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479

RESULT 2  
AX468543 1479 bp DNA linear PAT 16-JUL-2002  
LOCUS Sequence 63 from Patent WO0226209.  
DEFINITION AX468543  
ACCESSION AX468543  
VERSION AX468543.1 GI:21901373  
KEYWORDS  
SOURCE  
ORGANISM Human immunodeficiency virus 1 (HIV-1)  
REFERENCE Human immunodeficiency virus 1  
Virus; Retroviridae; Retroviridae; Lentivirus; Primate  
Lentivirus group.  
AUTHORS O'Hagan, D., Otten, G., Donnelly, J. J., Polo, J. M., Barnett, S.,  
Singh, M., Ulmer, J. and Dubensky, T. W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 63 04-Apr-2002;  
CHIRON CORPORATION (US)

FEATURES  
source location/Qualifiers  
1..1479  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"

ORIGIN

Query Match 100.0%; Score 1479; DB 6; Length 1479;  
Best Local Similarity 100.0%; Pred. No. 9.5e-151;  
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
DB 1 ATGGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
QY 61 CTGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 61 CTGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
QY 121 CTGGAAGAGTGGCTGTAACATGATGAAGCACTGTGTGTGGCCAGCGCGAG 180  
DB 121 CTGGAAGAGTGGCTGTAACATGATGAAGCACTGTGTGTGGCCAGCGCGAG 180  
QY 181 ATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 181 ATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
QY 241 ACCGTGGCACTGTGATCTGGTGAACGAGAAAGATGAGAGTCCGCAACCAAGAGGCC 300  
DB 241 ACCGTGGCACTGTGATCTGGTGAACGAGAAAGATGAGAGTCCGCAACCAAGAGGCC 300  
QY 301 CTGGAAGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 301 CTGGAAGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCCGCAAGAGGAGAGTGAAGCACTAATCCCATCTGTCAGAACTTGAGAGGCCAGATG 420  
DB 361 GCCGCAAGAGGAGAGTGAAGCACTAATCCCATCTGTCAGAACTTGAGAGGCCAGATG 420  
QY 421 GTGACACAGGCCATCAGCCCCCGCACTCTGAACGCTTGAGAGAGGCCAGGCCAG 480  
DB 421 GTGACACAGGCCATCAGCCCCCGCACTCTGAACGCTTGAGAGAGGCCAGGCCAG 480

QY 481 GCGTTCAGCCCGAGGTGATCCCAATGTTACCGCCCTGAGGAGGCGGCCAAGCCCGCAG 540  
DB 481 GCGTTCAGCCCGAGGTGATCCCAATGTTACCGCCCTGAGGAGGCGGCCAAGCCCGCAG 540  
QY 541 GACCTGAACACGATGTTGAACAACCGTGGCGGCGCCACAGCGCGCCATGAGATGCTGAAG 600  
DB 541 GACCTGAACACGATGTTGAACAACCGTGGCGGCGCCACAGCGCGCCATGAGATGCTGAAG 600  
QY 601 GACACATCAACAGAGAGGCGCCCGAGTGGACCCGCTGACACCCGCTGACCGCCGCGCC 660  
DB 601 GACACATCAACAGAGAGGCGCCCGAGTGGACCCGCTGACACCCGCTGACCGCCGCGCC 660  
QY 661 ATCGCCCCCGGAGATGCGAGGCGCGGAGGAGATGCGCGGAGCAACAGCAAGCAGC 720  
DB 661 ATCGCCCCCGGAGATGCGAGGCGCGGAGGAGATGCGCGGAGCAACAGCAAGCAGC 720  
QY 721 CTGCAAGAGAGATGCGCTGATGACCAAGCAACCCCGCATCCGCTGGGCGACATCTAC 780  
DB 721 CTGCAAGAGAGATGCGCTGATGACCAAGCAACCCCGCATCCGCTGGGCGACATCTAC 780  
QY 781 AAGCGGTGATCATCTGGGCTGAAACAAGATGTCGGATGTACAGCCCGTGAAGCATC 840  
DB 781 AAGCGGTGATCATCTGGGCTGAAACAAGATGTCGGATGTACAGCCCGTGAAGCATC 840  
QY 841 CTGCAATCAAGAGAGGCGCCCAAGAGGCGCTTCCGAGACTACGTGACCGCTTTTCAAG 900  
DB 841 CTGCAATCAAGAGAGGCGCCCAAGAGGCGCTTCCGAGACTACGTGACCGCTTTTCAAG 900  
QY 901 ACCCTGCGCGCGAGACAGACACCAAGAGTGAAGAACTGATGACAGCAACCTGCTG 960  
DB 901 ACCCTGCGCGCGAGACAGACACCAAGAGTGAAGAACTGATGACAGCAACCTGCTG 960  
QY 961 GTGCAGAAAGCCCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCGCGCGCGCAGC 1020  
DB 961 GTGCAGAAAGCCCAACCCCGACTGCAAGACCATCTGCGGCTCTCGGCGCGCGCGCAGC 1020  
QY 1021 CTGAGAGAGATGATACCGGCTGCAAGGCGCGGCGCGCGCGCGCGCGCGCGCGCG 1080  
DB 1021 CTGAGAGAGATGATACCGGCTGCAAGGCGCGGCGCGCGCGCGCGCGCGCGCGCG 1080  
QY 1081 CTGAGCGAGAGATGAGCCAGGCGCAACCAAGCTGATGATGACAGAGAGCACTTCAAG 1140  
DB 1081 CTGAGCGAGAGATGAGCCAGGCGCAACCAAGCTGATGATGACAGAGAGCACTTCAAG 1140  
QY 1141 GCGCCCGCGCGCATCTGCAAGTCTTCAACTGCGGCAAGAGGCGCCACATGCGCGCAAC 1200  
DB 1141 GCGCCCGCGCGCATCTGCAAGTCTTCAACTGCGGCAAGAGGCGCCACATGCGCGCAAC 1200  
QY 1201 TGCGCGCGCGCGCGAGAGAGGCGCTGAGAGTCCGCGCAAGAGGCGCCACAGATGAAG 1260  
DB 1201 TGCGCGCGCGCGCGAGAGAGGCGCTGAGAGTCCGCGCAAGAGGCGCCACAGATGAAG 1260  
QY 1261 GACTGACCGAGCGCGAGGCGCAACTTCTGAGGCAAGATCTGCGCGAGCGCAAGGAGCGCG 1320  
DB 1261 GACTGACCGAGCGCGAGGCGCAACTTCTGAGGCAAGATCTGCGCGAGCGCAAGGAGCGCG 1320  
QY 1321 CCGCGCAACTTCTGAGAGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
DB 1321 CCGCGCAACTTCTGAGAGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
QY 1381 TTTCAGAGAGACCAACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1381 TTTCAGAGAGACCAACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
QY 1441 CTGAAGAGCGTGTTCGCAACGAGCCCTGAGCGCACTAA 1479  
DB 1441 CTGAAGAGCGTGTTCGCAACGAGCCCTGAGCGCACTAA 1479

RESULT 3  
AX468547  
LOCUS AX468547 1479 bp DNA linear PAT 16-JUL-2002

DEFINITION Sequence 67 from Patent WO0226209.  
ACCESSION AX468547  
VERSION 1  
KEYWORDS GI:21901377  
SOURCE Human immunodeficiency virus 1 (HIV-1)  
ORGANISM Human immunodeficiency virus 1  
VIRUSES; Retroviruses; Retroviridae; Lentivirus; Primate  
Lentivirus group.  
REFERENCE 1  
AUTHORS O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
Singh,M., Umer,J. and Dubensky,T.W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 67 04-Apr-2002;  
CHIRON CORPORATION (US)  
FEATURES  
source location/Qualifiers  
1..1479  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"  
ORIGIN  
Query Match 99.6%; Score 1473.4; DB 6; Length 1479;  
Best Local Similarity 99.3%; Pred. No. 3.8e-150;  
Matches 1468; Conservative 10; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
DB 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
DB 61 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
QY 121 CTGAGAAAGTTCGCGCTGAAACCGCGCTCTGAGAACCGAGGAGGCGCTGCAAGCATC 180  
DB 121 CTGAGAAAGTTCGCGCGCTGAAACCGCGCTCTGAGAACCGAGGAGGCGCTGCAAGCATC 180  
QY 181 ATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
DB 181 ATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
QY 241 ACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
DB 241 ACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
QY 301 CTGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 301 CTGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
DB 361 GCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
QY 421 GTGCACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
DB 421 GTGCACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
QY 481 GCGTTCAGCCCGAGGTGATCCCAATGTTACCGCCCTGAGGAGGCGGCCAAGCCCGCAG 540  
DB 481 GCGTTCAGCCCGAGGTGATCCCAATGTTACCGCCCTGAGGAGGCGGCCAAGCCCGCAG 540  
QY 541 GACCTGAACACGATGTTGAACAACCGTGGCGGCGCCACAGCGCGCCATGAGATGCTGAAG 600  
DB 541 GACCTGAACACGATGTTGAACAACCGTGGCGGCGCCACAGCGCGCCATGAGATGCTGAAG 600  
QY 601 GACACATCAACAGAGAGGCGCGCGAGTGGACCCGCTGACACCCGCTGACCGCCGCGCC 660  
DB 601 GACACATCAACAGAGAGGCGCGCGAGTGGACCCGCTGACACCCGCTGACCGCCGCGCC 660  
QY 661 ATCGCCCCCGGAGATGCGAGGCGCGGAGGAGATGCGCGGAGCAACAGCAAGCAGCAGC 720  
DB 661 ATCGCCCCCGGAGATGCGAGGCGCGGAGGAGATGCGCGGAGCAACAGCAAGCAGCAGC 720

QY 721 CTGAGAGAGATGCGCTGATGACAGCAACCCCGGATCCCGGCGAGCATCTAC 780  
DB 721 CTGAGAGAGATGCGCTGATGACAGCAACCCCGGATCCCGGCGAGCATCTAC 780  
QY 781 AAGCGGTGATCATCTGGGCTGAAACAAGATGTCGGGATGTACAGCCCGTGAATC 840  
DB 781 AAGCGGTGATCATCTGGGCTGAAACAAGATGTCGGGATGTACAGCCCGTGAATC 840  
QY 841 CTGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
DB 841 CTGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 GTGAGAAAGCCCAACCCCGAGCTGCAAGACATCTGGGCTCTGGGCGCGAGC 1020  
DB 961 GTGAGAAAGCCCAACCCCGAGCTGCAAGACATCTGGGCTCTGGGCGCGAGC 1020  
QY 1021 CTGAGAGAGATGATGACCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
DB 1021 CTGAGAGAGATGATGACCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1081 CTGAGAGAGATGATGACCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
DB 1081 CTGAGAGAGATGATGACCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1141 GAGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
DB 1141 GAGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
QY 1201 TGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
DB 1201 TGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1261 GACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1261 GACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
QY 1321 CCGGCGAGATCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
DB 1321 CCGGCGAGATCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
QY 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
QY 1441 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
DB 1441 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479

RESULT 4  
AX455904 1479 bp DNA linear PART 06-JUL-2002  
LOCUS AX455904  
DEFINITION Sequence 20 from Patent WO0204493.  
ACCESSION AX455904  
VERSION AX455904.1 GI:21714896  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
synthetic construct  
other sequences; artificial sequences.  
REFERENCE  
1 zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
AUTHORS  
TITLE  
Polynucleotides encoding antigenic hiv type c polypeptides,  
Patent: WO 0204493-A 20 17-JAN-2002;  
JOURNAL  
CHIRON CORPORATION (US) : University of Stellenbosch (ZA)  
FEATURES  
Source  
1. 1479  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 98.9%; Score 1463; DB 6; Length 1479;  
Best Local Similarity 99.3%; Pred. No. 5e-149;  
Matches 1469; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
/db\_xref="taxon:32630"  
/note="Synthetic Gag coding sequence of HIV strain  
AP110965"

QY 1 ATGGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
DB 1 ATGGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
QY 61 CTGGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 61 CTGGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
QY 121 CTGAGAAAGTTGCGCTGTAACCCCGGCTGCTGAGAGAGAGAGAGAGAGAGAG 180  
DB 121 CTGAGAAAGTTGCGCTGTAACCCCGGCTGCTGAGAGAGAGAGAGAGAGAGAG 180  
QY 181 ATCGGCGAGCTGACACCCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 181 ATCGGCGAGCTGACACCCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
QY 241 ACCGTGGCGAGCTGATCTGGGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
DB 241 ACCGTGGCGAGCTGATCTGGGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
QY 301 CTGAGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 301 CTGAGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
DB 361 GCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
QY 421 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
DB 421 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
QY 481 GCTTTCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
DB 481 GCTTTCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
QY 541 GACTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
DB 541 GACTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
QY 601 GACACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 601 GACACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 661 ATGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
DB 661 ATGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 721 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
DB 721 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
QY 781 AAGCGGTGATCATCTGGGCTGAAACAAGATGTCGGGATGTACAGCCCGTGAATC 840  
DB 781 AAGCGGTGATCATCTGGGCTGAAACAAGATGTCGGGATGTACAGCCCGTGAATC 840  
QY 841 CTGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
DB 841 CTGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960



[illegible]

RESULT 5					
LOCUS	AY181195				
DEFINITION	AY181195	1485 bp	DNA	linear	SYN 02-OCT-2003
ACCESSION	AY181195				
VERSION					
KEYWORDS	AY181195.1	GI:37413992			
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	Gao,F., Li,Y., Decker,J.M., Payerl,F.W., Bibollet-Ruche,F., Rodenburg,C.M., Chen,Y., Shaw,D.R., Allen,S., Musonda,R., Shaw,G.M., Zafac,A.J., Letvin,N. and Hahn,B.H.				
TITLE	Codon usage optimization of HIV type 1 subtype C gag, pol, and nef genes: in vitro expression and immune responses in DNA-vaccinated mice				
JOURNAL	AIDS Res. Hum. Retroviruses	19 (9),	817-823	(2003)	
REFERENCE	2 (bases 1 to 1485)				
AUTHORS	Gao,F., Li,Y., Decker,J.M., Bibollet-Ruche,F., Rodenburg,C.M., Chen,Y., Shaw,D.R., Allen,S., Musonda,R., Shaw,G.M., Zafac,A.J. and Hahn,B.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-NOV-2002) Medicine,				
FEATURES	112 Refill, LaSalle Street, Durham,				
	NC 27710, USA				
	Location/Qualifiers				

```

source
1. .1485
   /organism="synthetic construct"
   /mol_type="genomic DNA"
   /db_xref="taxon:32630"
   /note="codon usage optimization of HIV-1 962M651"

gene
1. .1485
   /gene="gag"

CDS
1. .1485
   /gene="gag"

```

Query Match	89.8%	Score 1328.2	DB 12	Length 1485
Beat Local Similarity 94.3%	Pred. No. 1.7e-134			
Matches 1401	Conservative 0	Mismatches 78	Indels 6	Gaps 2
QY	1 ATGGGCGCGCGCCAGCATCTTCGCGCGCGGCAAGCTGACCGCTGGAGCGCATCCG	60		
DB	1 ATGGGCGCGCGCGCCAGCATCTTCGCGCGCGGCAAGCTGACCAATGGAGAAATCCG	60		
QY	61 CTGGCGCGCGCGCGCAAGAGTCTATCATATGAACACTGTGTGTGGCCACGCGCGAG	120		
DB	61 CTGGCGCGCGCGCGCAAGAGTCTATCATATGAACACTGTGTGTGGCCACGCGCGAG	120		
QY	121 CTGGAGAAAGTTGCGCCCTGGAACCCCGCGCTGTGGAGACAGCGAGGCTGCAACATC	180		
DB	121 CTGGAGGCGCTTCGCGCTCTGAACCCCGCGCTGTGGAGACAGCGAGGCTGCAACATC	180		
QY	181 ATCCGCCAGCTGCACCCCGCGCTTCAGACCGCGCGAGAGACTGAAGGCTGTTCAAC	240		
DB	181 ATGAAGAGCTGCACCCCGCGCTTCAGACCGCGCGAGAGACTGCGCAGCTGTCAAC	240		
QY	241 ACCGTGGCCACCTCTGTACTGTGCTACAGAGAAATCGAGTTCGCGGACACCAAGAGGCC	300		
DB	241 ACCGTGGCCACCTCTGTACTGTGCTACAGAGAGGCGCTGAGGGTGCACACCAAGAGGCC	300		
QY	301 CTGGAACAAGATCGAGAGAGAGCAGAAACAAGTGCAGAGAAATCCAGCAG--GCCGAG	357		
DB	301 CTGGAACGATCGAGAGAGAGCAGAAACAAGTGCAGAGAAATCCAGCAAAAACCGAG	360		
QY	358 GCCGCCGACAGAGGCAAGTGAAGCAGAACTACCCATCTGTCAGAACTTGCAGAGGCCAG	417		
DB	361 CAGGCCGCCGACAGGCAAGTGAAGCAGAACTACCCATCTGTCAGAACTTGCAGAGGCCAG	420		
QY	418 ATGTGTGACCAAGGCGCATCAGCCCGCGCACCTCTGAACGCTTGGGTGAAGTGTATGAGAG	477		
DB	421 ATGTGTGACCAAGAGCTGAGCCCGCGCACCTCTGAACGCTTGGGTGAAGTGTATGAGAG	480		
QY	478 AAGGCGTTACGCGCGCGAGTATATCCCATATTCAACGCGCTGAGCGAGGCCGCAACCCC	537		
DB	481 AAGGCGTTACGCGCGCGAGTATATCCCATATTCAACGCGCTGAGCGAGGCCGCAACCCC	540		
QY	538 CAGGACCTGAAACAGATGTTGAACAACGTTGGGCGGCAACAGGCGCGCATGAGATGCTG	597		
DB	541 CAGGACCTGAAACAGATGTTGAACAACGTTGGGCGGCAACAGGCGCGCATGAGATGCTG	600		
QY	598 AAGGACCAATCAACAGAGAGGCCCGGAGTGGAGCCGCGTGAACCCCTGTGACGCGGCG	657		
DB	601 AAGGACCAATCAACAGAGAGGCCCGGAGTGGAGCCGCGTGAACCCCTGTGACGCGGCG	660		
QY	658 CCATATGCGCGCGCGCGAGTGGCGGAGCCCGCGGAGCGACATCGCGCGGACCAACAGC	717		
DB	661 CCATATGCGCGCGCGCGAGTGGCGGAGCCCGCGGAGCGACATCGCGCGGACCAACAGC	720		
QY	718 ACCCTGAGAGAGCAGATCGCTGGATGACCAAGAACCCGCCCATCCCTGTGGGCGCATC	777		
DB	721 ACCCTGAGAGAGCAGATCGCTGGATGACCAAGAACCCGCCCATCCCTGTGGGCGCATC	780		
QY	778 TACAAGCGGTGATCATCTGGGCGTGAACAAGATCGTGGATGTACAAGCCCGGTGAGC	837		

Db 781 TACAGCCCTGATCATCTCTGGCCCTGAAAGATCGCGCATGTACAGCCCTGTAGC 840  
Qy 838 ATCTGACATCAAGCAGAGGCCCCCAAGAGGCCCTTCCCGCATCTAGTGAACGGCTTCTTC 897  
Db 841 ATCTGACATCAAGCAGAGGCCCCCAAGAGGCCCTTCCCGCATCTAGTGAACGGCTTCTTC 900  
Qy 898 AAGACCTGCGCGCGAGCAGAGCAACCCAGAGGTGAAGAACTGTATGACCGACACCTTG 957  
Db 901 AAGACCTGCGCGCGAGCAGAGCAACCCAGAGGTGAAGAACTGTATGACCGACACCTTG 960  
Qy 958 CTGTGAGGAAGCGCAACCCCGCATCTGCAAGACCATCTGCGCGCTCTGCGCCCGCGCGC 1017  
Db 961 CTGTGAGGAAGCGCAACCCCGCATCTGCAAGACCATCTGAGAGCCCTGCGCCCGCGCGC 1020  
Qy 1018 AGCCTGAGAGATGATGATACCGCTCTGCAAGGCGCTGAGGCGCCCGACCAAGGCCCGC 1077  
Db 1021 ACCCTGAGAGATGATGATACCGCTCTGCAAGGCGCTGAGGCGCCCGACCAAGGCCCGC 1080  
Qy 1078 GTGCTGCGGAGCGATGAGCAGGCGCAACAC--AGCTGATGATGAGAGAGCAAC 1134  
Db 1081 GTGCTGCGGAGCGATGAGCAGGCGCAACACAGCTGAAATCTCTGATGAGAGAGCAAC 1140  
Qy 1135 TTCAAGGGCCCCCGCGCATGCTCAAGTCTTCACTGCGGCAAGAGGGCGCATCGCC 1194  
Db 1141 TTCAAGGGCGAACAACCGCATGCTGAGTCTTCACTGCGGCAAGAGGGCGCATCGCC 1200  
Qy 1195 CGCAATCGCGCGCGCCCGCGCAAGAGGCGCTGCTGAAAGTGGCGCAAGAGGCCCAAG 1254  
Db 1201 CGCAATCGCGCGCGCCCGCGCAAGAGGCGCTGCTGAAAGTGGCGCAAGAGGCCCAAG 1260  
Qy 1255 ATGAAGACTGACCGAGCGCGCCCACTTCTCTGGGCAAGATCTGGCCCAAGCCCAAG 1314  
Db 1261 ATGAAGACTGACCGAGCGCGCCCACTTCTCTGGGCAAGATCTGGCCCAAGCCCAAG 1320  
Qy 1315 GCGCGCGCGCGCACTTCTCTGAGAGCGCGCCCGAGGCCACCGCCCGCGCGAGAGC 1374  
Db 1321 GCGCGCGCGCGCACTTCTCTGAGAGCGCGCCCGAGGCCACCGCCCGCGCGAGAGC 1380  
Qy 1375 TTCCGCTTGAAGAGCAACCCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434  
Db 1381 TTCCGCTTGAAGAGCAACCCCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Qy 1435 ACCAGCCTGAAGAGCTGTTGGGCAAGACCCCTCTGAGGCAAGTAA 1479  
Db 1441 ACCAGCCTGAAGAGCTGTTGGGCAAGACCCCTCTGAGGCAAGTAA 1485

RESULT 6  
AX455983  
LOCUS AX455983 1491 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 99 from Patent WO0204493.  
ACCESSION AX455983  
VERSION AX455983.1 GI:21714967  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
synthetic construct  
other sequences; artificial sequences.  
REFERENCE  
1 zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.  
AUTHORS Polynucleotide encoding antigenic h1v type c polypeptides,  
TITLE polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 99 17-JAN-2002;  
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
FEATURES  
source  
1..1491  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="tag\_IV2\_C\_ZaopL"

Query Match 87.1%; Score 1288.8; DB 6; Length 1491;

Best Local Similarity 93.0%; Pred. No. 3,1e-130;  
Matches 1389; Conservative 0; Mismatches 87; Indels 18; Gaps 3;  
Qy 1 ATGGCGCGCGCGCGCAGCATCTCTGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCG 60  
Db 1 ATGGCGCGCGCGCGCAGCATCTCTGCGCGCGCAAGCTGAGCAAGTGGAGAAATCCG 60  
Qy 61 CTGCGCGCGCGCGCGCAAGAAATGCTTACATGATGAAGCACTGTGTGTGGCGCAGCGAG 120  
Db 61 CTGCGCGCGCGCGCGCGCAAGCACTTACATGATGAGCACTGTGTGTGGCGCAGCGAG 120  
Qy 121 CTGGAAGATTCGCGCTGAAACCCCGCGCTGCGAGAGCAAGAGGCGCTGCAAGCATATC 180  
Db 121 CTGGAAGATTCGCGCGCTGAAACCCCGCGCTGCGAGAGCAAGAGGCGCTGCGCATATC 180  
Qy 181 ATTCGCAAGCTGCAACCCCGCTGCAAGCGCGCAGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 181 ATTCGCAAGCTGCAACCCCGCTGCAAGCGCGCAGAGAGAGAGAGAGAGAGAGAGAG 240  
Qy 241 ACCGTGCGCAACCTCTGTACTGTGTGTGCAAGAGAGATCGAGGTCCGCGCAACCAAGAG 300  
Db 241 ACCGTGCGCAACCTCTGTACTGTGTGTGCAAGAGAGATCGAGGTCCGCGCAACCAAGAG 300  
Qy 301 CTGGAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 301 CTGGAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Qy 361 GCGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Db 361 GCGCAACAAG--AAGTGTAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 417  
Qy 421 GTGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Db 421 GTGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Qy 418 GTGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 477  
Db 418 GTGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 477  
Qy 481 GCTTTAGCGCGCGAGAGATCCCATGTTCAACCGCGCTGAGAGAGAGAGAGAGAGAG 540  
Db 478 GCTTTAGCGCGCGAGAGATCCCATGTTCAACCGCGCTGAGAGAGAGAGAGAGAGAG 537  
Qy 541 GACTTGAACACATGTTGAACAACCGTGTGGCGCGCAACAGGCGCGCATATGCAATCTGAG 600  
Db 538 GACTTGAACACATGTTGAACAACCGTGTGGCGCGCAACAGGCGCGCATATGCAATCTGAG 597  
Qy 601 GACACCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Db 598 GACACCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657  
Qy 661 ATCGCGCGCGCGCAGATGCGAGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
Db 658 ATCGCGCGCGCGCAGATGCGAGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 717  
Qy 721 CTGCAAGAGAGAGATGCGCTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
Db 718 CTGCAAGAGAGAGATGCGCTGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 777  
Qy 781 AAGCGGTGATCATCTGTGGCGCTGAAACAAGATCTGCGAGATGTACAAGCGCGGTGAGCATC 840  
Db 778 AAGCGGTGATCATCTGTGGCGCTGAAACAAGATCTGCGAGATGTACAAGCGCGGTGAGCATC 837  
Qy 841 CTGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 838 CTGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897  
Qy 901 ACCGTGCGCGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 898 ACCGTGCGCGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957  
Qy 961 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 958 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017  
Qy 1021 CTGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080

Db	1018	CTGAGAGAGATGATATACCGCTCGCCAGGGCGGTGGGGGGCCCCAGGCCAAGAGGCCGGCTG	10177
Oy	1081	CTGGCCGAGGCGATGAGCCAGGC---CAACACACAGCGTATGATGACAGAGAGCACTTC	1137
Db	1078	CTGGCCGAGGCGCATGATGAGCCAGGCCAACAACACAGCGTATGATGACAGAGAGCACTTC	1137
Oy	1138	AAGGGCCCCCGGCGCATGTCTCAAGTGTCTTCAACTGCGGCAGAGAGGGCCACATGGCCGC	1197
Db	1138	AAGGGCCCCCGGCGCGCTGTGAAGTGTCTTCAACTGCGGCAGAGAGGGCCACATGGCCGC	1197
Oy	1198	AACTGCGCGCGCCCCCGGCAAGAGGGCTGTGTGAAGTGTGGCGAGAGAGGGCCACCATGT	1257
Db	1198	AACTGCGCGCGCCCCCGGCAAGGCGGCTGTGTGAAGTGTGGCAAGAGAGGGCCACCATGT	1257
Oy	1258	AAGGACTGACACCGAGCGGCCAGGCCCACTTCTCTGGCGAAGTCTTGGCCCAAGCCACAAGGGC	1317
Db	1258	AAGGACTGACACCGAGCGGCCAGGCCCACTTCTCTGGCGAAGTCTTGGCCCAAGCCACAAGGGC	1317
Oy	1318	CGCCCCGGGCAACTTCTCTGACGAGCGCGCCCGAGGCCACCGCCCCCGCGCGCGAGAGCTTC	1377
Db	1318	CGCCCCGGGCAACTTCTCTGACGAGCGCGCCCGAGGCCACCGCCCCCGCGCGCGAGAGCTTC	1377
Oy	1378	CGCTTTCGAGG-----AGACACACCCCGGCGCAAGACAGAGAGCAAGGACCGC	1425
Db	1378	GCCCCCCCCCGAGAGCTTCAAGTTCAAGAGACCCCCCAAGACAGAGACCCCAAGGACCGC	1437
Oy	1426	GAGACCTGACACAGGCTGAGAGGCTGTTCGGGCAAGACCCCGTGAAGCAGTAA	1479
Db	1438	GAGCCCCCTGACACAGCTGAGAGGCTGTTCGGGCAAGACCCCGTGAAGCAGTAA	1491

AX455935	LOCUS	AX455935	1494 bp	DNA	linear	PAT 06-JUL-2002
AX455935	DEFINITION	Sequence 51 from Patent WO0204493.				
AX455935	ACCESSION	AX455935				
AX455935.1	VERSION	GI:21714919				
	KEYWORDS					
SOURCE						
ORGANISM						
		synthetic construct				
		synthetic construct				
		other sequences; artificial sequences.				
REFERENCE						
AUTHORS		zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.				
TITLE		Polynucleotides encoding chimeric hiv type c polypeptides,				
		polypeptides and uses thereof				
		Patent: WO 0204493-A 51 17-JAN-2002;				
JOURNAL		CHIRON CORPORATION (US) : University of Stellenbosch (ZA)				
FEATURES						
Source						
		1..1494				
		/organism="synthetic construct"				
		/mol_type="unassigned DNA"				
		/db_xref="taxon:32630"				
		/note="HIV Type C Gag optimized"				
ORIGIN						
Query Match		87.1%;	Score 1288;	DB 6;	Length 1494;	
Best Local Similarity		92.9%;	Pred. No. 3.7e-130;			
Matches 1393;		Conservative 0;	Mismatches 80;	Indels 27;	Gaps 3;	
OY	1	ATGGGCGCCCGCGCAGCAGCATCTCGGCGCGCGGCAAGCTGGAACGCTGGAGAGGCAATCCGC	60			
Db	1	ATGGGCGCCCGCGCGCAGCAGCATCTGAGCGCGCGGCAAGCTGGAACGAGAGGCAATCCGC	60			
OY	61	CTGCGCGCCCGCGCGGCAAGAGTGCTACATGATGAAGCACTTGTTGGGCGCAGCCGCGAG	120			
Db	61	CTGCGCGCCCGCGCGGCAAGAGCACTACATGCTGAAGCACTGTGTGGCGCAGCGCGAG	120			
OY	121	CTGAGAGAGTTTCGCCCTGAACCCCGCGCTGCTGGAGACACGAGAGGCTGCAAGCAAGATC	180			
Db	121	CTGAGAGCCTTGGCCCTTAACCCCGCGCTGTGGAGACACGAGAGGCTGCAAGCAAGATC	180			
OY	181	ATCCGCAAGCTGACACCCGCGCTGGAAGCCGGAACGAGAGCTGAAGAGCCTGTTCAAC	240			

Db	181	ATCAAGCCTGACGCGCGCCCTGTGACACCGGCAACGAGACTGTGGCAGCTGTTCAC	24
OY	241	ACCGTGGCCACCTCTGTATCTGGGTGCACGAGAAATGATGAGGTCCGGACACCAAGAGGCG	300
Db	241	ACCGTGGCCACCTCTGTATCTGGGTGCACAAAGGCACTGAGGTGCGGACACCAAGAGGCG	300
OY	301	CTGGAACAAGATCGAGGAGAGACAGAAACAAGTGCAGAGAAATCAGACGGCCGAGGCG	360
Db	301	CTGGAACAAGATCGAGGAGAGACAGAAACAAGTGCAGAGAAATCAGACGGCCGAGGCG	360
OY	361	GCCGACAAGGCGCAAGGTAGGCGACAGAACTTACCCCATCTGTGACAACTGTGAGGCGCAGATG	420
Db	361	GCCGACAAGGCGCAAGGTAGGCGACAGAACTTACCCCATCTGTGACAAATGCCAGGCGCAGATG	417
OY	421	GTGCACCAAGCCATCAGCCCGCCGACCCCTTGAACCGCTTGGTGAAGGTGATCGAGAGAA	480
Db	418	GTGCACCAAGCCATCAGCCCGCCGACCCCTTGAACCGCTTGAATCAAGTGTATCGAGAGAA	477
OY	481	GCCTTCAAGCCCCGAGGTATATCCCATGTTTCAACCGGCTTGAAGGAGGCGCCACCCCGCAG	540
Db	478	GCCTTCAACCCCGAGGAGATCCCATGTTTCAACCGGCTTGAAGGAGGCGCCACCCCGCAG	537
OY	541	GACCTGAACACGATTTGTAAACAACGTGGAGCGGCGACAGAGCGCCATGCGATGCTGAAG	600
Db	538	GACCTGAACACATCTGTAAACAACGTGGAGCGGCGACAGAGCGCCATGCAAGATCTGAAG	597
OY	601	GACACCATCAACGAGAGAGCGCGCGAGTGGACCGCGTGCACCCCGTGCACGCGCGCGCC	660
Db	598	GACACCATCAACGAGAGAGCGCGCGAGTGGACCGCACCCCATCCCGTGCACGCGCGCGCC	657
OY	661	ATCGCCCCCGGCGAGATGCGCGAGGCCCCGCGGACGACGATGCGCGGACCAACAGACAC	720
Db	658	GTGGCCCCCGGCGAGATGCGCGAGGCCCCGCGGACGACGATGCGCGGACCAACAGACAC	717
OY	721	CTGCAGAGGAGGATGCGCTGTGATGACAGACAACCCCGCATCCCGTGGGCGACATCTAC	780
Db	718	CTGCAGAGGAGGATGCGCTGTGATGACAGACAACCCCGCATCCCGTGAAGACATCTAC	777
OY	781	AAGCGGTGATCATCTCTGGGCTTGAACAAGATCGTGCAGATGTACAGCCCGTGAACATC	840
Db	778	AAGCGGTGATCATCTCTGGGCTTGAACAAGATCGTGCAGATGTACAGCCCGTGAACATC	837
OY	841	CTGCACATCAAGCAGAGGCGCCCAAGAGGCTCTTCCGCGACTACGTGAGACCGCTTCTTCAAG	900
Db	838	CTGCACATCAAGCAGAGGCGCCCAAGAGGCTCTTCCGCGACTACGTGAGACCGCTTCTTCAAG	897
OY	901	ACCTTGGCGCGCGGACAGACACCCAGAGGTGAAGAACTGATATGACCGACACCCCTGCTG	960
Db	898	ACCTTGGCGCGCGGACAGACACCCAGAGGTGAAGAACTGATATGACCGACACCCCTGCTG	957
OY	961	GTGCAGAACGCGCAACCCCGGACTGCAAGACCAATCTGCGCGCTCTGGCGCCGCGGCGCAC	102
Db	958	GTGCAGAACGCGCAACCCCGGACTGCAAGACCAATCTGCGCGCTCTGGCGCCGCGGCGCAC	101
OY	1021	CTGAGAGAGATGATGACCGCCTTGCAGGAGCGTGGCGGCGCCGACCAAGGCGCGGCTG	108
Db	1018	CTGAGAGAGATGATGACCGCCTTGCAGGAGCGTGGCGGCGCCGACCAAGGCGCGGCTG	107
OY	1081	CTGGCCGAGCGGAGTGAAGCAGGCGCAACACAGCGTGAATGATGACAGAGCAACTTCAAG	114
Db	1078	CTGGCCGAGCGGAGTGAAGCAGGCGCAACACAGCGTGAATGATGACAGAGCAACTTCAAG	113
OY	1141	GGCCCCGGCGCATCTGTCAAGTGCTTCAACTGCGGCAAGAGGGGCAACTGCGCGGCAAC	120
Db	1138	GGCAGCAACCGCATCTATCAAGTGCTTCAACTGCGGCAAGAGGGGCAACTGCGCGGCAAC	119
OY	1201	TGCCGCGCGCCCCCGCAGAAAGGAGGCTGTGGAAGTGCAGCAAGAGGCGCCACCAATGAAG	126
Db	1198	TGCCGCGCGCCCCCGCAGAAAGGAGGCTGTGGAAGTGCAGCAAGAGGCGCCACCAATGAAG	125
OY	1261	GACTTGACATGAGCGCGCAGGCGCACTTCTTGGGCAAGATCTGGCCAGCCCAAGAGGCGCG	132
Db	1258	GACTTGACATGAGCGCGCAGGCGCACTTCTTGGGCAAGATCTGGGCGCAGCCCAAGAGGCGCG	131

Query Match	86.3%;	Score 1276.8;	DB 6;	Length 1509;
Best Local Similarity	92.5%;	Pred. No. 6e-129;		
Matches 1399;	Conservative	0;	Mismatches 77;	Indels 36; Gaps 4;
ORIGIN				
LOCUS	AX455888	1509 bp	DNA	linear
DEFINITION	Sequence 4 from Patent WO0204493.			
ACCESSION	AX455888			
VERSION	AX455888.1	GI:21714881		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE	1			
AUTHORS	zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.			
TITLE	Polynucleotides encoding antigenic hiv type c polypeptides, polypeptides and uses thereof			
JOURNAL	Patent: WO 0204493-A 4 17-JAN-2002;			
FEATURES	CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)			
SOURCE	1. 1509 /organism="synthetic construct" /mol_type="unassigned DNA" /db_xref="taxon:32630" /note="synthetic gag of HIV strain Afi10967"			
ORIGIN				
1	ATGGGCGCCCGCGCAGCATCTCTGGCGGCGGCAAGCTGGAAGCGCTGGAGGCGATCCGC	60		
1	ATGGGCGCCCGCGCAGCATCTCTGGCGGCGGCAAGCTGGAAGCGCTGGAGGCGATCCGC	60		
61	CTGCGCGCCCGCGGCGGAAGAGTGCTTACATGATGAAGCACTTGCTGTGAGCGCAGCCGCGAG	120		
61	CTGCGCGCCCGCGGCGGAAGAGTGCTTACATGATGAAGCACTTGCTGTGAGCGCAGCCGCGAG	120		
61	CTGCGCGCCCGCGGCGGAAGAGTGCTTACATGATGAAGCACTTGCTGTGAGCGCAGCCGCGAG	120		
121	CTGGAGGAAGTTGCGCCCTGAACCCCGCGCTGCTGGAGACACGGAAGGCTGCAAGCAAGATC	180		
121	CTGGAGGAAGTTGCGCCCTGAACCCCGCGCTGCTGGAGACACGGAAGGCTGCAAGCAAGATC	180		
121	CTGGAGGAAGTTGCGCCCTGAACCCCGCGCTGCTGGAGACACGGAAGGCTGCAAGCAAGATC	180		
181	ATCCCGCAGCTGCAACCCCGCCTCTGGAGACCGGCAACGGAAGAGCTGAAAGCCTGTTCAAC	240		
181	ATCCCGCAGCTGCAACCCCGCCTCTGGAGACCGGCAACGGAAGAGCTGAAAGCCTGTTCAAC	240		
181	ATCCCGCAGCTGCAACCCCGCCTCTGGAGACCGGCAACGGAAGAGCTGAAAGCCTGTTCAAC	240		
241	ACCGTGGCCACCTCTGTACTGCTGACCGCGGCATCTGAGGTCCTCGACACCAAGAGGCGC	300		
241	ACCGTGGCCACCTCTGTACTGCTGACCGCGGCATCTGAGGTCCTCGACACCAAGAGGCGC	300		
241	ACCGTGGCCACCTCTGTACTGCTGACCGCGGCATCTGAGGTCCTCGACACCAAGAGGCGC	300		
301	CTGGACAAGATGAGAGAGAGAGCAAGATCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	360		
301	CTGGACAAGATGAGAGAGAGAGCAAGATCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	360		
301	CTGGACAAGATGAGAGAGAGAGCAAGATCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	360		
361	GCCGCAAGAGGCGCAAGGTGAGCGCAAGTACCCCATCTGTGCAAGACCTGCAAGGCGCAGATG	420		
361	GCCGCAAGAGGCGCAAGGTGAGCGCAAGTACCCCATCTGTGCAAGACCTGCAAGGCGCAGATG	420		
361	GCCGCAAGAGGCGCAAGGTGAGCGCAAGTACCCCATCTGTGCAAGACCTGCAAGGCGCAGATG	420		
421	GTGCAACCAAGGCGCATAGCCCGCGACCTGAACTGCTGCTGGAGAGAGTATCGAGAGAGAG	480		
421	GTGCAACCAAGGCGCATAGCCCGCGACCTGAACTGCTGCTGGAGAGAGTATCGAGAGAGAG	480		
421	GTGCAACCAAGGCGCATAGCCCGCGACCTGAACTGCTGCTGGAGAGAGTATCGAGAGAGAG	480		
418	GTGCAACCAAGGCGCATAGCCCGCGACCTGAACTGCTGCTGGAGAGAGTATCGAGAGAGAG	477		
418	GTGCAACCAAGGCGCATAGCCCGCGACCTGAACTGCTGCTGGAGAGAGTATCGAGAGAGAG	477		
418	GTGCAACCAAGGCGCATAGCCCGCGACCTGAACTGCTGCTGGAGAGAGTATCGAGAGAGAG	477		

QY	441	GCTTTACGCCCCAGAGTATCCCATGTTTCAACCGCCCTGAGGAGAGGCGCCACCCCCGAC	540
Db	478	GCTTTACGCCCCGAGGTATCCCATGTTTCAACCGCCCTGAGGAGAGGCGCCACCCCCGAC	537
QY	541	GACCTGAACAAGATTGTGAACAACCGTGGCGCGCCACAAGGCGCGCATGCAGATGTGAAG	600
Db	538	GACCTGAACAAGATTGTGAACAACCGTGGCGCGCCACAAGGCGCGCATGCAGATGTGAAG	597
QY	601	GACACCATCAACGAGAGAGGCGCGCGAGTGGGAACCGGTGACACCCGTGTGACCGCGGCCCC	660
Db	598	GACACCATCAACGAGAGAGGCGCGCGAGTGGGAACCGGTGTGACACCCGTGTGAGCGCGGCCCC	657
QY	661	ATCGCCCCCGGCATATGCGGAGGCCCCGCGGAGGAGCATGCGCGGCAACAAGACACC	720
Db	658	GTGCCCCCGGCATATGCGGAGGCCCCGCGGAGGAGCATGCGCGGCAACAAGACACC	717
QY	721	CTGCAGAGAGCAGATGCGCTGTGATGAACAACAACCCCCCATCCCGTGGGCGACATCTAC	780
Db	718	CTGCAGAGAGCAGATGCGCTGTGATGAACAACAACCCCCCATCCCGTGGGCGACATCTAC	777
QY	781	AAGCGGTGATATATCTTGCGGCTTGAAACAAGATCTGTGCGGATGTTACAGCCCCGTGAGCAAC	840
Db	778	AAGCGGTGATATATCTTGCGGCTTGAAACAAGATCTGTGCGGATGTTACAGCCCCGTGAGCAATC	837
QY	841	CTGCAGATCAAGCAGAGGCCCCCAAGAGGCGCTTCCGCGACTACGTTGAGACCGCTTCTTCAAG	900
Db	838	CTGCAGATCTCCGCAAGGCCCCCAAGAGGCGCTTCCGCGACTACGTTGAGACCGCTTCTTCAAG	897
QY	901	ACCCTGCGCGCGGAGCAGACACCAAGAGGTGAAGAACTGGATGACGACACCTGTCTG	960
Db	898	ACCCTGCGCGCGGAGCAGGCGCAACCAAGAGGTGAAGAACTGGATGACGACACCTGTCTG	957
QY	961	GTGCGAAGACGCCAACCCCGGACTGCAAGAACATCTCGCGCGCTCTGCGGCCCGGCGCGAC	1021
Db	958	GTGCGAAGACGCCAACCCCGGACTGCAAGAACATCTCGCGCGCTCTGCGGCCCGGCGCGAC	1011
QY	1021	CTGAGAGAGATGATGATGACCGGCTTGCCAGAGGCGTGTGGCGGCGCCAGCCACAAGGCGCGGTG	1081
Db	1018	CTGAGAGAGATGATGATGACCGGCTTGCCAGAGGCGTGTGGCGGCGCCAGCCACAAGGCGCGGTG	1071
QY	1081	CTGCGCGGAGGCGATGAGCCAGGCGCAACAC---ACCTGATGATGACGAAGAAGCACTTCC	1131
Db	1078	CTGCGCGGAGGCGATGAGCCAGGCGCAACACGCTGAACATCATGATGACGAAGAAGCACTTCC	1131
QY	1138	AAGGCGCCCCGCGCGATGCTCAAGTGTCTTCAACTGCGGCAAGAGGCGCCACATCGCGCGC	1191
Db	1138	AAGGCGCCCCGCGCGCATGCTCAAGTGTCTTCAACTGCGGCAAGAGGCGCCACATCGCGCAAG	1191
QY	1198	AACTGCGCGCGCCCCCGCAAGAGAGGCTGTGTAAGTGTGGGCAAGAGGCGCCACCAAGATG	1251
Db	1198	AACTGCGCGCGCCCCCGCAAGAGAGGCTGTGTAAGTGTGGGCAAGAGGCGCCACCAAGATG	1251
QY	1258	AAGGACTGGAACCGAGCGCGGCGGCAACTCTCTGGGCAAGATCTGGGCGCAAGCCACAAGGCG	1311
Db	1258	AAGGACTGGAACCGAGCGCGGCGGCAACTCTCTGGGCAAGATCTGGGCGCAAGCCACAAGGCG	1311
QY	1318	CGCCCGCGGCACTTCTCTGCAAGCGG-----CCCGAGCCCAACCGCGC	1351
Db	1318	CGCCCGCGGCACTTCTCTGCAAGACCGGAGCCCGCGCGCCCGCCACCGTGCACACGCGC	1371
QY	1360	CCCCCGCGGAGAGCTTCCGCTTTCGAGGAGACACCCCGGCGCAAGAGCAAGAGAGCAAG	1411
Db	1378	CCCCCGCGGAGAGCTTCCGCTTTCGAGGAGACACCCCGGCGCAAGAGAGAGAGCCCAAG	1431
QY	1420	G-----ACCGGAGAGCCCTGACCAAGCTTGAAGAGGCTGTTGGGCAACGACCGCC	1461
Db	1438	GACCGGAGCCCTTACCGGAGGCGCTGACCGGCGCTGTGCGAGGCTGTTCCGAGCGGCGCCC	1491
QY	1468	CTGAGCCAGTAA	1479
Db	1498	CTGAGCCAGTAA	1509



AUTHORS O'Hagan,D., Otten,G., Donnelly,J.J., Polo,J.M., Barnett,S.,  
Singh,M., Ulmer,J. and Dubensky,T.W.  
TITLE Microparticles for delivery of the heterologous nucleic acids  
JOURNAL Patent: WO 0226209-A 68 04-Apr-2002;  
CHIRON CORPORATION (US)  
FEATURES Location/Qualifiers  
source 1. 1509  
/organism="Human immunodeficiency virus 1"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:11676"

## ORIGIN

Query Match 86.1%; Score 1273.6; DB 6; Length 1509;  
Best Local Similarity 91.8%; Pred. No. 1.3e-128;  
Matches 1388; Conservative 12; Mismatches 76; Indels 36; Gaps 4;

1 ATGGGCGCCCGCGCAGATCCTGCGCGCGCAGATGACGCTGAGAGCGCATCCGC 60  
1 ATGGGCGCCCGCGCAGATCCTGCGCGCGCAGATGACGATGGGAGAGATCCGC 60  
61 CTGGCGCCCGCGCGCAGAGATGATGATGAGACCTGCTGTGGGCGCAGCCGCGAG 120  
61 CTGGCGCCCGCGCGCAGAGATGATGATGAGACCTGCTGTGGGCGCAGCCGCGAG 120  
121 CTGGAAGATGCTGCGCTGAAACCGCGCGCTGCTGAGACCAAGGAGGCTGCAAGCATC 180  
121 CTGGAAGATGCTGCGCTGAAACCGCGCGCTGCTGAGACCAAGGAGGCTGCAAGCATC 180  
181 ATCCGCGAGCTGCAACCGCGCGCTGCGAGACCGCGCAGAGAGCTGAGAGCTGTTCAAC 240  
181 ATGGAAGAGCTGCAACCGCGCGCTGCGAGACCGCGCAGAGAGCTGAGAGCTGTTCAAC 240  
241 ACCGTGCGACCTCTGTAATGCTGTGCAAGAGAGATGAGTCCGCGCAACCAAGAGGCC 300  
241 ACCGTGCGACCTCTGTAATGCTGTGCAAGAGAGATGAGTCCGCGCAACCAAGAGGCC 300  
301 CTGGAAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
301 CTGGAAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
361 GCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
361 GCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
421 GTGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
421 GTGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
481 GCTTCACGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
481 GCTTCACGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
541 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
541 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
538 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597  
538 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597  
601 GACACATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
601 GACACATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
598 GACACATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657  
598 GACACATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657  
661 ATGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
661 ATGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
658 GTGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717  
658 GTGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717  
721 CTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
721 CTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
718 CTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 777  
718 CTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 777  
781 AAGCGGTGATCATCTGCGGCTTGAACAAGATGCTGCGAGATTAAGCCCGGTGAGCATC 840  
781 AAGCGGTGATCATCTGCGGCTTGAACAAGATGCTGCGAGATTAAGCCCGGTGAGCATC 840  
778 AAGCGGTGATCATCTGCGGCTTGAACAAGATGCTGCGAGATTAAGCCCGGTGAGCATC 837  
778 AAGCGGTGATCATCTGCGGCTTGAACAAGATGCTGCGAGATTAAGCCCGGTGAGCATC 837  
841 CTGCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
841 CTGCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900

DB 838 CTGCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897  
QY 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 898 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957  
QY 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
DB 958 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017  
QY 1021 CTGGAAGAGATGATGACCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
DB 1018 CTGGAAGAGATGATGACCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077  
QY 1081 CTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137  
DB 1078 CTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137  
QY 1138 AAGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197  
DB 1138 AAGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197  
QY 1198 AACTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
DB 1198 AACTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
QY 1258 AAGGAGCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317  
DB 1258 AAGGAGCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317  
QY 1318 CGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1359  
DB 1318 CGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1359  
QY 1360 CCCCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1419  
DB 1378 CCCCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
QY 1420 G-----ACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1467  
DB 1438 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
QY 1468 CTGAGCGAGTAA 1479  
DB 1498 CTGAGCGAGTAA 1509

RESULT 11  
AX455905 1509 bp DNA linear PAT 06-JUL-2002  
LOCUS AX455905  
DEFINITION Sequence 21 from Patent WO204493.  
ACCESSION AX455905  
VERSION AX455905.1 GI:21714897  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
other sequences; artificial sequences.

REFERENCE  
1  
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Renenburg,E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 21 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)  
FEATURES Location/Qualifiers  
source 1. 1509  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="synthetic Gag coding sequence of HIV strain AF110967"

## ORIGIN

Query Match 85.2%; Score 1260.8; DB 6; Length 1509;  
 Best Local Similarity 91.1%; Pred. No. 3.2e-127;  
 Matches 1389; Conservative 0; Mismatches 87; Indels 36; Gaps 4;

```

QY 1 ATGGGCGCCGCGGCGAGATCTTGGCGGCGGCAAGTGAAGCTTGGAGCGCATCCGC 60
DB 1 ATGGGCGCCGCGGCGAGATCTTGGCGGCGGCAAGTGAAGCTTGGAGCGCATCCGC 60
QY 61 CTGGCGCCCGCGGCGAAGAGTGTACATGATGAAGCACTGTGTGTGGCCGCGCGAG 120
DB 61 CTGGCGCCCGCGGCGAAGAGTGTACATGATGAAGCACTGTGTGTGGCCGCGCGAG 120
QY 121 CTGGAGAGTTCGCGCTTGAACCCCGGCTGTGTGAGACGAGGAGGCTGCAAGAGATC 180
DB 121 CTGGAGAGTTCGCGCTTGAACCCCGGCTGTGTGAGACGAGGAGGCTGCAAGAGATC 180
QY 181 ATCCGCGCGTGCACCCCGCGCTGTGAGACGCGGACGAGAGCTGAGAGCTGTTCAC 240
DB 181 ATGAAGGAGCTGCAGCCGCGCTGTGAGACGCGGACGAGAGCTGAGAGCTGTTCAC 240
QY 241 ACCGTGGCGACCTCTGTACTGTGTGACGAGAAAGTGTGCGGACACCAAGAGGCG 300
DB 241 ACCGTGGCGACCTCTGTACTGTGTGACGAGAAAGTGTGCGGACACCAAGAGGCG 300
QY 301 CTGGACAGATCGAGGAGAGCAAGCAAGTGCAGAGAGATCAGAGGCGGAGGCG 360
DB 301 CTGGACAGATCGAGGAGAGCAAGCAAGTGCAGAGAGATCAGAGGCGGAGGCG 360
QY 361 GCCGCAAGGCGAGAGTGAAGCAGAACTAACCCCATGTGCAAGACCTGAGGCGCAGATG 420
DB 361 GCCGCAAGGCGAGAGTGAAGCAGAACTAACCCCATGTGCAAGACCTGAGGCGCAGATG 420
QY 421 GTGCAACGAGGCGATAGCCCGCGCACCTGAAAGCTGTGGTGAAGTGTGAGAGAG 480
DB 421 GTGCAACGAGGCGATAGCCCGCGCACCTGAAAGCTGTGGTGAAGTGTGAGAGAG 480
QY 481 GCGTTACGCGCGAGGATGCCCATGTTCAACCGGCTGAGAGGAGGCGGCGGCGGCGAG 540
DB 481 GCGTTACGCGCGAGGATGCCCATGTTCAACCGGCTGAGAGGAGGCGGCGGCGGCGAG 540
QY 541 GACCTGAACACGATTTGAAACACGCTGGGCGGCGCACAGGCGCGCATGTGCTGAAG 600
DB 541 GACCTGAACACGATTTGAAACACGCTGGGCGGCGCACAGGCGCGCATGTGCTGAAG 600
QY 597 GACCTGAACACGATTTGAAACACGCTGGGCGGCGCACAGGCGCGCATGTGCTGAAG 597
DB 597 GACCTGAACACGATTTGAAACACGCTGGGCGGCGCACAGGCGCGCATGTGCTGAAG 597
QY 601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCGTGCACCCGTGACGCGCGCGCC 660
DB 601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCGTGCACCCGTGACGCGCGCGCC 660
QY 657 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCGTGCACCCGTGACGCGCGCGCC 657
DB 657 GACACCATCAACGAGAGAGCGCGCGAGTGGAGCCGCGTGCACCCGTGACGCGCGCGCC 657
QY 720 ATCCGCGCGCGGCGAGTGCAGAGCGCGCGGCGAGCATGTGCGGCGCACCAAGCGCACCC 720
DB 720 ATCCGCGCGCGGCGAGTGCAGAGCGCGCGGCGAGCATGTGCGGCGCACCAAGCGCACCC 720
QY 717 GTGGCGCGCGGCGAGATGCGCACCCCGGCGGCGAGCATGTGCGGCGCACCAAGCGCACCC 717
DB 717 GTGGCGCGCGGCGAGATGCGCACCCCGGCGGCGAGCATGTGCGGCGCACCAAGCGCACCC 717
QY 780 CTGGAGAGAGAGATGCGCTGTGATGACAGCAACCCCGCATCCCGTGGGCGACATCTAC 780
DB 780 CTGGAGAGAGAGATGCGCTGTGATGACAGCAACCCCGCATCCCGTGGGCGACATCTAC 780
QY 777 CTGGAGAGAGAGATGCGCTGTGATGACAGCAACCCCGCATCCCGTGGGCGACATCTAC 777
DB 777 CTGGAGAGAGAGATGCGCTGTGATGACAGCAACCCCGCATCCCGTGGGCGACATCTAC 777
QY 840 AAGCGGTGATGATCTTGGGCTGAAACAAGATGTGCGAGATGTACAGCCCGGTGAGCATC 840
DB 840 AAGCGGTGATGATCTTGGGCTGAAACAAGATGTGCGAGATGTACAGCCCGGTGAGCATC 840
QY 837 AAGCGGTGATGATCTTGGGCTGAAACAAGATGTGCGAGATGTACAGCCCGGTGAGCATC 837
DB 837 AAGCGGTGATGATCTTGGGCTGAAACAAGATGTGCGAGATGTACAGCCCGGTGAGCATC 837
QY 900 CTGGACATCAAGCAAGGCGCGCAAGAGCGCTTCCGCGACTAGTGAACCGCTTCTTCAAG 900
DB 900 CTGGACATCAAGCAAGGCGCGCAAGAGCGCTTCCGCGACTAGTGAACCGCTTCTTCAAG 900
QY 897 CTGGACATCAAGCAAGGCGCGCAAGAGCGCTTCCGCGACTAGTGAACCGCTTCTTCAAG 897
DB 897 CTGGACATCAAGCAAGGCGCGCAAGAGCGCTTCCGCGACTAGTGAACCGCTTCTTCAAG 897
QY 960 ACCCTGCGCGCGGCGAGAGCAAGCAAGAGGTGAAGACTGTGATGACAGCAACCTGTGCTG 960
DB 960 ACCCTGCGCGCGGCGAGAGCAAGCAAGAGGTGAAGACTGTGATGACAGCAACCTGTGCTG 960
QY 957 ACCCTGCGCGCGGCGAGAGCAAGCAAGAGGTGAAGACTGTGATGACAGCAACCTGTGCTG 957
DB 957 ACCCTGCGCGCGGCGAGAGCAAGCAAGAGGTGAAGACTGTGATGACAGCAACCTGTGCTG 957
QY 1020 GTGCGAGAAAGCGCAACCTCGGACTGCAAGACCATCTTGGCGGCTCTGCGCGCGCGCACG 1020
DB 1020 GTGCGAGAAAGCGCAACCTCGGACTGCAAGACCATCTTGGCGGCTCTGCGCGCGCGCACG 1020
QY 1017 GTGCGAGAAAGCGCAACCTCGGACTGCAAGACCATCTTGGCGGCGCTCTGCGCGCGCACG 1017
DB 1017 GTGCGAGAAAGCGCAACCTCGGACTGCAAGACCATCTTGGCGGCGCTCTGCGCGCGCACG 1017
QY 1080 CTGGAGAGAGATGATACCGCGCTGCAAGGCGGTGGCGGCGCGCCACGACCAAGGCGCGCGT 1080
DB 1080 CTGGAGAGAGATGATACCGCGCTGCAAGGCGGTGGCGGCGCGCGCCACGACCAAGGCGCGCGT 1080

```

Query Match 82.6%; Score 1221.2; DB 6; Length 9166;  
 Best Local Similarity 90.1%; Pred. No. 3.5e-123;  
 Matches 1346; Conservative 0; Mismatches 133; Indels 15; Gaps 3;

```

DB 1018 CTGGAGAGAGATGATACCGCGCTGCGAGGCGCGTGGCGCGCGCGCGCGCGCGCG 1077
QY 1081 CTGGCGCGGCGGATGATGACGAGCGCAACAC---AGCGTATGATGACAGAAAGCAATTC 1137
DB 1078 CTGGCGCGGCGGATGATGACGAGCGCAACAC---AGCGTATGATGATGATGATGATGATG 1137
QY 1138 AAGGCGCGCGCGGCGATGTCAAGTGTTCATCTGCGGCAAGAGAGGCGCAATCGCGCG 1197
DB 1138 AAGGCGCGCGCGGCGATGTCAAGTGTTCATCTGCGGCAAGAGAGGCGCAATCGCGCG 1197
QY 1198 AATGCGCGCGCGCGCGCGGCGGCAAGAGGCTGTGAAAGTGTGCGGCAAGAGGCGCAAGATG 1257
DB 1198 AATGCGCGCGCGCGCGCGGCGGCAAGAGGCTGTGAAAGTGTGCGGCAAGAGGCGCAAGATG 1257
QY 1258 AAGGACTGCAACCGAGCGCGGCGCAACTTCTGTGGCAAGATGTGGCCAGCGCAAGAGGCG 1317
DB 1258 AAGGACTGCAACCGAGCGCGGCGCAACTTCTGTGGCAAGATGTGGCCAGCGCAAGAGGCG 1317
QY 1318 GCGCGCGCGCAACTTCTGTGAGAGCGG-----CGCGAGCGCGCGCGCGCGCGCGCG 1359
DB 1318 GCGCGCGCGCAACTTCTGTGAGAGCGGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377
QY 1360 CCGCGCGCGGAGAGCTTCTGAGAGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1419
DB 1378 CCGCGCGCGGAGAGCTTCTGAGAGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437
QY 1420 G-----ACCGGAGACCTGTGACAGGCTGAAAGAGCTGTGGCGACGACCGCG 1467
DB 1438 GACCGGAGCGCGCTTCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1497
QY 1468 CTGAGCGCAGTAA 1479
DB 1498 CTGAGCGCAGTAA 1509

RESULT 12
AX427930
LOCUS AX427930 9166 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 168 from Patent WO232943.
ACCESSION AX427930
VERSION AX427930.1 GI:21538017
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Huang, Y. and Nabel, G.J.
Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
Patent: WO 0232943-A 168 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
1. 9166
/oranism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/feature="Plasmid pVR1012x/b containing HIV genes"
ORIGIN
Query Match 82.6%; Score 1221.2; DB 6; Length 9166;  

  Best Local Similarity 90.1%; Pred. No. 3.5e-123;  

  Matches 1346; Conservative 0; Mismatches 133; Indels 15; Gaps 3;


```

QY 1 ATGGGCGCGCGGCGAGATCTTGGCGGCGGCAAGTGAAGCTTGGAGCGCATCCGC 60
DB 1880 ATGGGCGCGCGGCGAGATCTTGGCGGCGGCAAGTGAAGCTTGGAGCGCATCCGC 1939
QY 61 CTGGCGCCCGCGGCGAAGAGTGTACATGATGAAGCACTGTGTGTGGCCGCGCGAG 120
DB 1940 CTGAAGCGCGGCGGCAAGAGCACTATGATGAAGCACTGTGTGTGGCCGCGCGAG 1999
QY 121 CTGGAGAGTTCGCGCTTGAACCCCGGCTGTGTGAGACGAGGAGGCTGCAAGAGATC 180

```


```



Db 2000 CTGAGAGCCCTTGCCTTGAGACCCCGGCTGTGAGAGACCAAGAGGCTGCAAGAGATC 2055  
Qy 181 ATCCGCGAGTGCACCCCGCTTGAGACCCGAGCGAGAGCTGAAGACCTGTTCAAC 240  
Db 2060 ATGAAGCATGTGCAGCCCGCTTGAGACCCGCAACAGAGCTGATGACCTTGACAAC 2119  
Qy 241 ACCGCGCAGCCCTGTACTGTGTGACGAGAGATCGAGTCCCGGACACCAAGAGAGCC 300  
Db 2120 ACCGCGCAGCCCTGTACTGTGTGACGAGAGATCGAGTCCCGGACACCAAGAGAGCC 2179  
Qy 301 CTGACCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCC 360  
Db 2180 CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCC 2239  
Qy 361 GCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Db 2240 GCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2299  
Qy 421 GTGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Db 2300 GTGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2359  
Qy 481 GCCTTCAGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
Db 2360 GCCTTCAGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2419  
Qy 541 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
Db 2420 GACCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2479  
Qy 601 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Db 2480 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2539  
Qy 661 ATCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
Db 2540 ATCCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2599  
Qy 721 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
Db 2600 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2659  
Qy 781 AAGCGGTGAGATCATCTGAGGCTGAGAACAGAGATCGAGAGATGATGAGCCGTGAGCATC 840  
Db 2660 AAGCGGTGAGATCATCTGAGGCTGAGAACAGAGATCGAGAGATGATGAGCCGTGAGCATC 2719  
Qy 841 CTGAGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 2720 CTGAGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2779  
Qy 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 2780 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2839  
Qy 961 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 2840 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2899  
Qy 1021 CTGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
Db 2900 CTGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2959  
Qy 1081 CTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137  
Db 2960 CTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3019  
Qy 1138 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197  
Db 3020 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3079  
Qy 1198 AACTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257

Db 3080 AACTGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3139  
Qy 1258 AAGAGCTGCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317  
Db 3140 AAGAGCTGCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3199  
Qy 1318 CGCCCGCGCACTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377  
Db 3200 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3259  
Qy 1378 CGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425  
Db 3260 AGGTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3319  
Qy 1426 GAGACCTGCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
Db 3320 TATCTTTAGCTTCCTCAGATCACTCTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3373

RESULT 13  
AX427936  
LOCUS AX427936 9788 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 174 from Patent WO0232943.  
ACCESSION AX427936  
VERSION AX427936.1 GI:21538023  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
other sequences; artificial sequences.  
REFERENCE  
1 Huang, Y. and Nabel, G.J.  
Modifications of hiv env, gag, and pol enhance immunogenicity for  
genetic immunization  
Patent: WO 0232943-A 174 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
LOCATION/Qualifiers  
1..9788  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Plasmid pVR1012x/s containing HIV genes"

ORIGIN  
Query Match 81.5%; Score 1206; DB 6; Length 9788;  
Best Local Similarity 90.0%; Pred. No. 1.5e-121;  
Matches 1344; Conservative 0; Mismatches 130; Indels 20; Gaps 4;  
Qy 1 ATGAGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
Db 1880 ATGAGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1939  
Qy 61 CTGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
Db 1940 CTGAGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1999  
Qy 121 CTGAGAGAGTCCGCTGAGAACCGCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
Db 2000 CTGAGAGCGCTTGCCTGAGAACCGCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2059  
Qy 181 ATCCGAGAGTGCACCCCGCTTGAGACCCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 2060 ATGAAGAGAGTGCAGCCCGCTTGAGACCCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2119  
Qy 241 ACCGCGCAGCCCTGTACTGTGTGACGAGAGATCGAGTCCCGGACACCAAGAGAGCC 300  
Db 2120 ACCGCGCAGCCCTGTACTGTGTGACGAGAGATCGAGTCCCGGACACCAAGAGAGCC 2179  
Qy 301 CTGACCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCC 360  
Db 2180 CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCC 2239  
Qy 361 GCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420



```
Db 2240 GCCGACAAAGGAGGATGAGCCAGAACTACCCCATCTGTGACAGAACTCTGAGGAGCCAGATG 2299
Qy 421 GTGACACAGGCGATCAGCCCCCGGACCCCTGAAAGCTCTGGGTGAAGGTATCGAGAGAAAG 480
Db 2300 GTGACACAGGCGCATCAGCGCCCGCACCCCTGAAAGCTCTGGGTGAAGGTATCGAGAGAAAG 2359
Qy 481 GCGCTTCAGGCGGAGGTGATCCCAATGTTTCAACCGGCTTGAAGGAGGAGCCACCCCGGAG 540
Db 2380 GCGCTTCAGGCGGAGGTGATCCCAATGTTTCAACCGGCTTGAAGGAGGAGGAGCCACCCCGGAG 2419
Qy 541 GACCTGAACACGATGTTGAACAACCGTGGGCGGACACAGGCGGCAATGCAATGCTGAAG 600
Db 2420 GACCTGAACACCATGTCTGAACAACCGTGGGCGGACACAGGCGGCAATGCAATGCTGAAG 2479
Qy 601 GACACCATCAACAGAGAGGCGCGGAGTGGGACCGGCTGACACCCGCTGACCGCGGCGGCGG 660
Db 2480 GACACCATCAACAGAGAGGCGCGGAGTGGGACCGGCTGACACCCGCTGACCGCGGCGGCGG 2539
Qy 661 ATCGCCCCCGGAGATGCGGAGCGGAGCGGCGGAGGAGCATGCGCGGAGCAACCAAGCACC 720
Db 2540 GTGCGCCCCCGGAGATGCGGAGCGGAGCGGCGGAGGAGCATGCGCGGAGCAACCAAGCACC 2599
Qy 721 CTGCAAGAGCAGATCGCTGATGACACAGCAACCCCGCCATCCCGTGGGCGACATCTAC 780
Db 2600 CTGCAAGAGCAGATCACTGATGACCAACCAACCCCGCCATCCCGTGGGCGACATCTAC 2659
Qy 781 AAGCGGTGATCATCTGGGCTGAAACAAAGATCTGCGGATGTAACGCCCTGTAGCATC 840
Db 2660 AAGCGGTGATCATCTGGGCTGAAACAAAGATCTGCGGATGTAACGCCCTGTAGCATC 2719
Qy 841 CTGCAATCAACAGAGGCGGCGGAGGAGCGGCTTCCGCGACCTAGTGAACCGCTTCTTCAAG 900
Db 2720 CTGCAATCAACAGAGGCGGCGGAGGAGCGGCTTCCGCGACCTAGTGAACCGCTTCTTCAAG 2779
Qy 901 ACCCTGCGCGGAGCAGACAGCAACCAAGAGTGAAGAACTGATGACCGACACCTGCTG 960
Db 2780 ACCCTGCGCGGAGCAGACAGCAACCAAGAGTGAAGAACTGATGACCGACACCTGCTG 2839
Qy 961 GTGCAAAAGCCCAACCCGATCTGCAAGACCATCTGCGGCTCTGCGGCGGCGGCGGCGG 1020
Db 2840 GTGCAAAAGCCCAACCCGATCTGCAAGACCATCTGCGGCTCTGCGGCGGCGGCGGCGG 2899
Qy 1021 CTGAGAGGATGATGACCGGCTGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Db 2900 CTGAGAGGATGATGACCGGCTGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2959
Qy 1081 CTGAGCGAGGCGATGAGCGAGG---CAACACAGCGTGTATGATGACAGAGAGCAACTTC 1137
Db 2960 CTGAGCGAGGCGATGAGCGAGG---CAACACAGCGTGTATGATGACAGAGAGCAACTTC 3019
Qy 1138 AAGGCGCGGCGGCGATCTGCAAGTCTTCAACTGCGGCGGAGAGGCGGCGGCGGCGG 1197
Db 3020 AAGGCGCGGCGGCGATCTGCAAGTCTTCAACTGCGGCGGAGAGGCGGCGGCGGCGG 3079
Qy 1198 AACTGCGGCGGCGGCGGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1257
Db 3080 AACTGCGGCGGCGGCGGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3139
Qy 1258 AAGGATGACCGAGCGGCGGCGCAACTTCCTGCGGCAAGATCTGCGGCGGCGGCGGCGG 1317
Db 3140 AAGGATGACCGAGCGGCGGCGCAACTTCCTGCGGCAAGATCTGCGGCGGCGGCGGCGG 3194
Qy 1318 CGCGCGGCGGCACTTCTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1377
Db 3195 AGGCGGAGGAAATTTCTTCAAGAGAGACCAAGCGGCGGCGGCGGCGGCGGCGGCGG 3254
Qy 1378 CGCTTCGAGGAG-----GACCAACCGCGGCGGAGAGAGAGAG-----AGCAAGAGCGG 1425
Db 3255 AGGTTTGGGAGAGAGCAACAATCTCTCTTCAAGAGAGAGAGCGGCGGCGGCGGCGG 3314
Qy 1426 GAGACCTGACAGCGCTGAAAGAGCGTGTTCGCGCAAGAGCGGCGGCGGCGGCGGCGG 1479
Db 3315 TATCTTTAGCTTCTCTGATCAGTCTTTGGGAGAGAGCGGCGGCGGCGGCGGCGGCGG 3368
```

```
RESULT 14
AXI49648
LOCUS AXI49648 4288 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 2 from Patent WO0136614.
ACCESSION AXI49648
VERSION AXI49648.1 GI:14348047
KEYWORDS
SOURCE Human immunodeficiency virus
ORGANISM Human immunodeficiency virus
Virus; Retrovirus; Retroviridae; Lentivirus; Primate
lentivirus group.
REFERENCE
1 Shao, Y., Wagner, R., Wolf, H. and Graf, M.
AUTHORS The genome of the HIV-1 inter-subtype (c/b') and use thereof
JOURNAL Patent: WO 0136614-A 2 25-MAY-2001;
Geneart GmbH Gesellschaft fuer angewandte Biotechnologie (DE) ;
Shao, Yiming (CN)
FEATURES
location/Qualifiers
source 1..4288
/organism="Human immunodeficiency virus"
/mol_type="unassigned DNA"
/db_xref="taxon:12721"
ORIGIN
Query Match 81.5%; Score 1205.4; DB 6; Length 4288;
Best Local Similarity 89.5%; Pred. No. 2.2e-121;
Matches 1335; Conservative 0; Mismatches 141; Indels 15; Gaps 3;
Qy 1 ATGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
Db 13 ATGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 72
Qy 61 CTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 73 CTGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 132
Qy 121 CTGAGAAAGTTTGGCTTGAACCCCGGCTGCTGAGAGACCGAGGAGGCTGCAAGCATG 180
Db 133 CTGAGAAAGTTTGGCTTGAACCCCGGCTGCTGAGAGACCGAGGAGGCTGCAAGCATG 192
Qy 181 ATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 193 ATGAAGCGGCTGCAAGAGCGGCTTGCAGACCGGCAACGAGAGCTGAGAGCTTGTTCAC 252
Qy 241 ACCGTGCGCACCTGTACTGTGCTGCAAGAGATCGAGGTCGCGGACCAAGAGGCG 300
Db 253 ACCGTGCGCACCTGTACTGTGCTGCAAGAGATCGAGGTCGCGGACCAAGAGGCG 312
Qy 301 CTGAGCAAGATGAGAGAGAGAGAGCAAAAGTCCAGAGAGATCGAGAGGCGGCGGCGG 360
Db 313 CTGAGCAAGATGAGAGAGAGAGAGCAAAAGTCCAGAGAGATCGAGAGGCGGCGGCGG 372
Qy 361 GCGGCAAGGCGGCAAGTGAAGAGAGAACTAACCCATCGTGCAGAACTGAGAGGCGGCGG 420
Db 373 GCGGCAAGGCGGCAAGTGAAGAGAGAACTAACCCATCGTGCAGAACTGAGAGGCGGCGG 429
Qy 421 GTGCAACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 430 GTGCAACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 489
Qy 481 GCGTTGAGCGCGGAGGTGATCCCAATGTTTCAACCGGCTTGAAGGAGGCGGCGGCGGCGG 540
Db 490 GCGTTGAGCGCGGAGGTGATCCCAATGTTTCAACCGGCTTGAAGGAGGCGGCGGCGGCGG 549
Qy 541 GACCTGAACACGATGTTGAACAACCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Db 550 GACCTGAACACGATGTTGAACAACCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 609
Qy 601 GACACCATCAACAGAGAGGCGGCGGAGTGGGACCGGCTGACACCCGCTGACCGCGGCGG 660
Db 610 GACACCATCAACAGAGAGGCGGCGGAGTGGGACCGGCTGACACCCGCTGACCGCGGCGG 669
```

QY	661	ATTCGCCCCCGGCATATGCGGAGGCCCCCGGCGGACGACATATGCGGCGACCAACGACGAC	720
Db	670	ATTCGCCCCCGGCATATGCGGAGGCCCCCGGCGGACGACATATGCGGCGACCAACGACGAC	729
QY	721	CTGCAGAGACGATGCGCTGTGATGACAGCAACCCCCCATCCCGTGGCGACATCTTAC	780
Db	730	CTGCAGAGACGATGCGCTGTGATGACAGCAACCCCCCATCCCGTGGCGACATCTTAC	789
QY	781	AAGCGGTGATCATCTTGGGGCTTGAACAAAGATCGTGGGATGTACAGCCCGGTGAGCATC	840
Db	790	AAGAGGTGATCATCTTGGGTTTAAACAAGATCGTGGGATGTACAGCCCGGTGAGCATC	849
QY	841	CTGCACATCAAGCAGAGGCCCCCAAGAGAGCCCTTCGCGACCTACGTGACCGCTTCTTCAAG	900
Db	850	CTGCACATCAAGCAGAGGCCCCCAAGAGAGCCCTTCAGGAGCTACGTGACAGGTTCTTCAAG	909
QY	901	ACCTTCGCGCGCCGAGCAGACACCCAGAGAGGTGAAGAACTGAGTACCGACACCTCTGCTG	960
Db	910	ACCTGAGAGGCGGAGCAGGCGCACCCAGAGGCGTGAAGAACTGAGTACCGACACCTCTGCTG	969
QY	961	GTCGAGAAACGCCAACCCCGACCTTGCAGAACCATCTTGGCGGCTCTGGGCCCCGCGCCAGC	1022
Db	970	GTCGAGAAACGCCAACCCCGACCTTGCAGAACCATCTTGAAGGCCCTTGGGCCCCGCGCCAGC	1022
QY	1021	CTGCAGAGATGATGATACCGGCTTGCAGAGGCGGTGGGGCGGCGCCAGACCAAGAGCCGCGGTG	1080
Db	1030	ATTCAGAGAGATGATATCCGCTTGCAGAGGCGGTGGGGCGGCGCCAGACCAAGAGCGTGG	1088
QY	1081	CTGCAGAGCGCATGATGAGCGGACCAACACAGCGTATGATGACAGAGAGCAACTTCAAG	1140
Db	1090	CTGCAGAGCGCATGATGAGCGGACCAACACAGCGCATCTGATGACAGAGAGCAACTTCAAG	1144
QY	1141	GGCCCCCGGCGCATGTGTAAGTCTTCAACTTGGCGCAGAGAGGCGCATGTGCCCCGCAAC	1200
Db	1150	GCGACCAAGAGGATGTGTAAGTCTTCAACTTGGCGCAGAGAGGCGCATGTGCCCCGCAAG	1203
QY	1201	TGCCCGCGCCCCCGCAGAGAGGCGCTGTGAAAGTGTCGCGCAAGAGGCGCACAGATGTAG	1260
Db	1210	TGCAGAGGCCCCCGCAGAGAGAGGCGCTGTGAAAGTGTCGCGCAAGAGGCGCACAGATGTAG	1266
QY	1261	GACTGCACCGAGCGGCAGGCGCAACTTCTTGGGCAAGATCTTGGCCAGCGCCACAGAGGCGCG	1320
Db	1270	GACTGCACCGAGAGGCGAGGCGCAACTTCTTGGGCAAGATCTTGGCCAGCGCCACAGAGGCGCG	1322
QY	1321	CCCCGCACTTCTCTGACAGACCGGCCCCGACCGCCACCGCCCCCGCGGAGGCTTCCGC	1380
Db	1330	CCCCGCACTTCTCTGACAGACCGGCCCCGACCGCCACCGCCCCCGGAGGAGCTTCAAG	1389
QY	1381	TTTCGAGAG-----ACCAACCCCGCGCGCAGAGCAGAG-----AGCAAGAGCCGCGAG	1428
Db	1390	TTTCGAGAGAGAGACCAACCAACCCCGCGCAGAGAGAGAGGCCATGACCAAGAGGCTGTAC	1444
QY	1429	ACCTGCACCGAGCTTGAAGAGCTGTTCGAGCAAGACCCCTGAGCCGATGA 1479	
Db	1450	CCCTGCACCGAGCTTGAAGAGCTGTTCGAGCAAGACCCGAGCAGCCAGGAA 1500	
RESULT 15			
LOCUS	AF201927	1509 bp	DNA linear SYN 16-MAR-2000
DEFINITION	Synthetic construct gag protein gene, complete cds.		
ACCESSION	AF201927		
VERSION	AF201927.1	GI:7248702	
KEYWORDS	synthetic construct		
SOURCE	synthetic construct		
ORGANISM	other sequences; artificial sequences.		
REFERENCE	1 (bases 1 to 1509)		
AUTHORS	Zur Meede, J., Chen, M.C., Doe, B., Schaefer, M., Greer, C.E., Selby, M., Otten, G.R. and Barnett, S.W.		
TITLE	Increased expression and immunogenicity of sequence-modified human immunodeficiency virus type 1 gag gene		

JOURNAL U Virol. 74 (6), 2628-2635 (2000)  
 MEDLINE 20148954  
 PUBMED 10684277  
 REFERENCE 2 (bases 1 to 1509)  
 AUTHORS zur Megeide, J. and Barnette, S.W.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-NOV-1999) Vaccines, Chiron Corporation, 4560 Horton,  
 Emeryville, CA 94608, USA  
 FEATURES  
 source  
 location/Qualifiers  
 1..1509  
 /organism="synthetic construct"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32630"  
 1..1509  
 /note="HIV-1SP2 p55gag"  
 /codon\_start=1  
 /transl\_table=11  
 /product="gag protein"  
 /protein\_id="AA043628.1"  
 /db\_xref="GI:7248703"  
 /translation="MGARASVTSGGELDKWEKIRLPQGGKKRYKIKLHIVASRELER  
 AVNGSLTSEGGKQIIGOLPISLOTGSESLRYNTVAITVCHORIDVDTKALV  
 KIEBONKSKKKAOIAAAGCTGSSVSOVNPVYONOGOMVHOAISPRILNAVTK  
 VEEKARSPREVPMTSALSSEGTAPRODITMDITLVGHQAAOMKLTINEEASMDRVRV  
 PVHGPAPQGMREPRKSDIAGITSTLDEOIGMMTNPNPIFVGEIKYKMTILGLNKITY  
 RMYSPSTSLIDIRQGPKEPRFDYVDRFYKTLAEASQDVKNMTEITLVONAPDCKT  
 ILKILAPATLEEMMTACOGVGPGHKRYVLAEMSQTNNPATIMOGNPNRDKTY  
 KCPFGKEGHTFARNCPARKKGCWRKCGEHSQMDCTERQANFLGIKIPSYKGRGNIN  
 LQSRPEPTAPPEESFRFEKTKTTSQKQEPIDKELYPLTSLRSIFGNDPSSO"  
 ORIGIN  
 Query Match 81.0%; Score 1197.8; DB 12; Length 1509;  
 Best Local Similarity 89.3%; Pred. No. 2e-120;  
 Matches 1347; Conservative 0; Mismatches 132; Indels 30; Gaps 4  
 QY 1 ATGGGCGCGCGCGCGCAGCATCTCTGGCGGCGGCGCAAGCTGAGACGCTTGGAGCGCATCCGC 60  
 DB 1 ATGGGCGCGCGCGCGCAGCGTGTCTGACGCGCGCGCGCAGCTGAGCAAGTGGAGAAAGTATCCGC 60  
 QY 61 CTGGCGCGCGCGCGCGCAGAAAGTGTCTACTAGTAAAGCACTGTGTGGGCGACCGCGGAG 120  
 DB 61 CTGGCGCGCGCGCGCGCAAGAAAGTAAAGTAAAGTAAAGCACTGTGTGGGCGACCGCGGAG 120  
 QY 121 CTGGAGAAAGTTTCGGCTCTGAACCCCGGCGCTGTGGAGACCAAGCGAGCGCTGCAAGAGATC 180  
 DB 121 CTGGAGCGCTTCGCGGTGAACCCCGGCGCTGTGGAGACCAAGCGAGCGCTGCGCGAGATC 180  
 QY 181 ATCGCGCAGCTGACCCCGCGCTTGAGACCGGCGAGCGAGCGAGCTGMAAGCTGTCTTCAAC 240  
 DB 181 CTGGCGCAGCTGACCGCCAGCCTGAGACCGGCGAGCGAGCGAGCTGCGACCTGTAAAC 240  
 QY 241 ACCGTGGCCACCTCTGTATCTGGCTGTGACGAGAAAGTGAAGTTCGAGATCCGGAACAGAGGCC 300  
 DB 241 ACCGTGGCCACCTCTGTATCTGGCTGTGACGAGGCAATCGTCAAGAGAACCAAGAGAGGCC 300  
 QY 301 CTGGCACAAGATTCAGAGAGAGAGACAGAAACAAGTGCAGCGAAGATTCACACAGGCCCGAGGCC 360  
 DB 301 CTGGAGAAAGATTCAGAGAGAGAGACAGAAACAAGTTCAGAAAGAAAGGCCACAGGCCCGCGCC 360  
 QY 361 GCCG-----ACAAAGGCGAAGGTGAGCGCAAGAACTAACCCATGTGTGCAAACTTG 408  
 DB 361 GCCGCGGCAACCGGCAACAGCAACAGCAGCGAGTGAACCACTAACCCATGTGTGCAAACTTG 420  
 QY 409 CAGGGCCAGATGTGTGACCAAGGCCATCAAGCCCCCGCACCCCTGAAGAGCTTGGGTGAAGGTG 468  
 DB 421 CAGGGCCAGATGTGTGACCAAGGCCATCAAGCCCCCGCACCTGAAGAGCTTGGGTGAAGGTG 480  
 QY 469 ATCGAGAGAAAGGCTTCAGCGCCGAGGTGTATCCCATGTTCACCGCCTTGAGCGAGGAC 528  
 DB 481 GTGAGAGAGAAAGGCTTCAGCGCCGAGGTGTATCCCATGTTCAGCGCCTTGAGCGAGGAC 540  
 QY 529 GCCACCCCCAGAGACTTGAACAAGATTTTGAACACCGTGGGCGGCGCAACAGGCCCGCATG 588

```
Db 541 GCCACCCCCCAGGACCTGAACAGATGTTGAACACCGTGGCGGCCACCAAGGCCGCATG 600
QY 589 CAGATGCTGAAGACACCATCAACGAGAGGCCGCCGAGTGGGACCGCGTGCACCCCGTGG 648
Db 601 CAGATGCTGAAGAGACCATCAACGAGAGGCCGCCGAGTGGGACCGCGTGCACCCCGTGG 660
QY 649 CAGCGCGGCGCCCATCGCCCCCGGCGCAGATGCGCGAGAGCCCGCGCGCAGCGACATCGCGG 708
Db 661 CAGCGCGGCGCCCATCGCCCCCGGCGCAGATGCGCGAGAGCCCGCGCGCAGCGACATCGCGG 720
QY 709 ACCACGAGCACCCTGAGAGAGCATGCGCTGGATGACCAACCCGCCATCCCGCTGG 768
Db 721 ACCACGAGCACCCTGAGAGAGCATGCGCTGGATGACCAACCCGCCATCCCGCTGG 780
QY 769 GCGGACATTTACAGAGCGGTGGATCATCTGGGGCTTGAACAAGATCGTGGGATTTACAGC 828
Db 781 GCGGAGATCTACAGAGCGGTGGATCATCTGGGGCTTGAACAAGATCGTGGGATTTACAGC 840
QY 829 CCGGTGAGCATCTGGACATCAAGCAGGGGCCCAAGAGCCCTTCGCGCATACGTGAC 888
Db 841 CCGACCGACATCTGGACATCCGCCAGGGGCCCAAGAGCCCTTCGCGCATACGTGAC 900
QY 889 CGCTTCTTCAAGACCTGCGCGCCGAGCAGAGCACCAGAGGTGAAGAACTGATGACC 948
Db 901 CGCTTCTTCAAGACCTGCGCGCTGAGCAGGCGCAGAGCGTGAAGAACTGATGACC 960
QY 949 GACACCTGTGTGTGAGAACCGCAACCCCGACTGCAAGACCATCTGCGCGCTCTCGGC 1008
Db 961 GAGACCTGTGTGTGAGAACCGCAACCCCGACTGCAAGACCATCTGAAAGCTCTCGGC 1020
QY 1009 CCGGCGCGCAGCGTGGAGAGATGATGAGCGGCTGCGCAGGGCGTGGCGGCCCGACGCAC 1068
Db 1021 CCGGCGCGCAGCGTGGAGAGATGATGAGCGGCTGCGCAGGGCGTGGCGGCCCGACGCAC 1080
QY 1069 AAGGCCCGGCTGTGGCGCAGAGCGATGAGCCAGCCACA-----CCAGCGTATGATG 1122
Db 1081 AAGGCCCGGCTGTGGCGCAGAGCGATGAGCCAGCGTGAAGACCCGCGCAGCATCATGATG 1140
QY 1123 CAGAAAGCAATTCAAGGGCCCCCGGCGCATGTGTCAAGTGTTCATCTGCGCAAGAG 1182
Db 1141 CAGCGCGGCACTTCGCAACAGCGGAAAGACCGTCAAGTGTTCATCTGCGCAAGAG 1200
QY 1183 GGGCCATGCGCGCGCACTGCGCGCGCGCGCGCGCGCGCAAGAGGCGTGGAAATGCGGCAAG 1242
Db 1201 GGGCCATGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCAAGAGGCGTGGCGCGCGC 1260
QY 1243 GAGGCGCACCATGAGAGATGTCACCGAGCGCCAGGCCAATTCTTGGGCAAGATCTGG 1302
Db 1261 GAGGCGCACCATGAGAGATGTCACCGAGCGCGCGCAATTCTTGGGCAAGATCTGG 1320
QY 1303 CCGAGCCACAGAGGCGCGCGCGCGCAATTCTTGCAGAGCGCGCGCGCAACCGCGCGC 1362
Db 1321 CCGAGCTACAGAGGCGCGCGCGCGCAATTCTTGCAGAGCGCGCGCGCAACCGCGCGC 1380
QY 1363 CCGGCGAGAGGTTCCGCTTC-----GAGGAGACCAACCCCGCGCGCAAGAGAG--- 1413
Db 1381 CCGGAGAGAGGTTCCGCTTCGCGCGAGAGAGACCAACCCCGAGAGAGAGAGCGCC 1440
QY 1414 ---AGCAAGAGACCGGAGACCTTGACCAAGAGCGCTGTTGCGCAACGACCCCGTGG 1470
Db 1441 ATCGACAGAGAGCTGTACCCCTTGACCAAGCTGCGGAGCTGTTCGCGCAACGACCCAGC 1500
QY 1471 AGCCAGTAA 1479
Db 1501 AGCCAGTAA 1509
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 11:36:50 ; Search time 839.981 Seconds  
(without alignments)  
10423.210 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 1479

Sequence: 1 atggggcgcgcgcgcagcat.....acgacccctgagccagtaa 1479

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn19808:\*  
2: geneseqn19908:\*  
3: geneseqn20008:\*  
4: geneseqn20018:\*  
5: geneseqn20018:\*  
6: geneseqn20028:\*  
7: geneseqn20028:\*  
8: geneseqn20038:\*  
9: geneseqn20038:\*  
10: geneseqn20038:\*  
11: geneseqn20038:\*  
12: geneseqn20048:\*  
13: geneseqn20048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1479	100.0	1479	3	AAAS1609
2	1479	100.0	1479	6	AA144548
3	1479	100.0	1479	6	AB139953
4	1479	100.0	1479	12	ADM73758
5	1477.4	99.9	3162	10	ADCI13239
6	1477.4	99.9	3462	10	ADCI13237
7	1477.4	99.9	4419	8	ACA03523
8	1477.4	99.9	4419	10	ADCI13240
9	1477.4	99.9	4615	10	ADCI13257
10	1477.4	99.9	4702	10	ADCI13259
11	1477.4	99.9	4716	8	ACA03522
12	1477.4	99.9	4716	10	ADCI13238
13	1475.8	99.8	2742	8	ACA03524
14	1475.8	99.8	2742	10	ADCI13241
15	1475.8	99.8	3330	10	ADCI13230
16	1475.8	99.8	3330	10	ADCI13231
17	1475.8	99.8	3330	10	ADCI13232
18	1475.8	99.8	5145	8	ACA03521
19	1475.8	99.8	5145	10	ADCI13233
20	1474.8	99.7	4713	8	ACA03522

21	1474.8	99.7	4713	10	ADCI13280	Adci13280 DNA of HI
22	1473.8	99.6	5184	8	ACA03591	ACA03591 Synthetic
23	1473.8	99.6	5184	10	ADCI13279	Adci13279 DNA of HI
24	1472.4	99.6	2742	8	ACA03590	ACA03590 Synthetic
25	1472.4	99.6	2742	10	ADCI13278	Adci13278 DNA of HI
26	1463	98.9	1479	3	AAAS1625	AAAS1625 HIV codon
27	1463	98.9	1479	6	AB139957	Ab139957 Synthetic
28	1463	98.9	1479	12	ADM73762	Adm73762 HIV-1 pol
29	1461.4	98.8	1479	6	AA144552	AA144552 HIV-1 p55
30	1424.4	96.3	4546	10	ADCI13255	Adci13255 DNA of HI
31	1347.4	91.1	4423	10	ADCI13256	Adci13256 DNA of HI
32	1315.2	88.9	3531	10	ADCI13234	Adci13234 DNA of HI
33	1315.2	88.9	3537	10	ADCI13236	Adci13236 DNA of HI
34	1315.2	88.9	3538	10	ADCI13235	Adci13235 DNA of HI
35	1288.8	87.1	1491	6	AB140020	AB140020 Synthetic
36	1288.8	87.1	1491	12	ADM73833	Adm73833 HIV-1 pol
37	1288	87.1	1494	6	AB139972	Ab139972 Synthetic
38	1276.8	86.3	1509	3	AAAS1610	AAAS1610 HIV synth
39	1276.8	86.3	1509	6	AA144549	AA144549 HIV-1 p55
40	1276.8	86.3	1509	6	AB139954	Ab139954 Synthetic
41	1276.8	86.3	1509	6	AB139954	Ab139954 Synthetic
42	1276.8	86.3	1509	12	ADM73759	Adm73759 HIV-1 pol
43	1262.4	85.4	1509	6	AB139958	Ab139958 Synthetic
44	1262.4	85.4	1509	12	ADM73763	Adm73763 HIV-1 pol
45	1260.8	85.2	1509	3	AAAS1626	AAAS1626 HIV codon

## ALIGNMENTS

RESULT 1	AAAS1609	standard; DNA, 1479 BP.
ID	AAAS1609	
XX	AAAS1609;	
AC		
XX		
DT	31-OCT-2000	(first entry)
XX		
DE	HIV synthetic Gag polynucleotide.	
XX		
KW	Gag; expression cassette; antigenic; type C; HIV; Env; synthetic;	
XX	DNA immunization; packaging cell line; antigen presentation; ss.	
OS	Human immunodeficiency virus; type C strain AF110965.	
XX	Synthetic.	
FT	Key	Location/Qualifiers
FT	CDS	1..1479
FT		/*tag="a
FT		/product="Synthetic_Gag"
FT		/note="Codon usage pattern was modified and inhibitory
FT		elements (INS) and RRE sites were inactivated resulting
FT		in improved expression"
XX		
XX	WO200039304-A2.	
XX	06-JUL-2000.	
PD		
XX	30-DEC-1999;	99WO-US031273.
PF		
XX	31-DEC-1998;	98US-0114495P.
PR	01-SEP-1999;	99US-0152195P.
XX		
PA	(CHIR ) CHIRON CORP.	
XX		
PI	Barnett S, Zur Megede J;	
XX		
DR	WPI: 2000-452401/39.	
XX	F-PSDB; AA196943.	
DR		
XX		
PT	Polynucleotide encoding antigenic type C HIV Gag polypeptide or a HIV Env	
PT	polypeptide and the polypeptide useful for immunizing a mammal especially	
PT	human against HIV.	

XX Claim 2, Page 92-93; 113pp; English.  
 XX Expression cassettes comprising a polynucleotide encoding antigenic type  
 CC C human immunodeficiency virus (HIV) Gag or Env polypeptides are useful  
 CC in DNA immunization, generation of packaging cell lines and production of  
 CC Gag- and/or Env-containing proteins. Synthetic Env and Gag expression  
 CC cassettes exhibit increased potency for induction of cytotoxic T-  
 CC lymphocyte (CTL) responses by DNA immunization. Gag of HIV-1 self-  
 CC assemble into non-infectious virus-like particles which are used as a  
 CC matrix for the proper presentation of an antigen entrapped or associated  
 CC to the immune system of the host

XX Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;

Query Match 100.0%; Score 1479; DB 3; Length 1479;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-185;  
 Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGGGCGCGCGCGCATCTCTGGCGCGCGCGCAAGCTGAGCGCCCTGGAGGCGCATCCGC 60  
 DB 1 ATGGGGGCGCGCGCGCATCTCTGGCGCGCGCGCAAGCTGAGCGCCCTGGAGGCGCATCCGC 60  
 QY 61 CTGGCGCGCGCGCGCAAGAGTCTACATGATGAACTCTGGTGTGGCGCAAGCGCGAG 120  
 DB 61 CTGGCGCGCGCGCGCAAGAGTCTACATGATGAACTCTGGTGTGGCGCAAGCGCGAG 120  
 QY 121 CTGGAGAAATTGCTGCTGAACCCCGGCTGTGGAGACAGCGAGGGCTGCAAGCAGATC 180  
 DB 121 CTGGAGAAATTGCTGCTGAACCCCGGCTGTGGAGACAGCGAGGGCTGCAAGCAGATC 180  
 QY 181 ATCCGCGAGCTGACCCCGGCTGTGAGACCGGAGAGAGAGAGCTGAGAGGCTGTTCAC 240  
 DB 181 ATCCGCGAGCTGACCCCGGCTGTGAGACCGGAGAGAGAGAGCTGAGAGGCTGTTCAC 240  
 QY 241 ACCGTGGCCACCTGTACTGCTGCAAGAGATGAGTCCGCGACCAAGAGAGGC 300  
 DB 241 ACCGTGGCCACCTGTACTGCTGCAAGAGATGAGTCCGCGACCAAGAGAGGC 300  
 QY 301 CTGGACAGAGATGAGAGAGAGAGCAAAAGTCCAGAGAGATCCAGAGCGCC 360  
 DB 301 CTGGACAGAGATGAGAGAGAGAGCAAAAGTCCAGAGAGATCCAGAGCGCC 360  
 QY 361 GCCGACAGAGAGAGAGTGAAGCACTACCCCATGTGTCAGAACTTGAGGGCCAGATG 420  
 DB 361 GCCGACAGAGAGAGAGTGAAGCACTACCCCATGTGTCAGAACTTGAGGGCCAGATG 420  
 QY 421 GTGACACAGAGGCATAGCCCGCGACCTGTAAAGCTGTGGTGAAGTGAAGAGAG 480  
 DB 421 GTGACACAGAGGCATAGCCCGCGACCTGTAAAGCTGTGGTGAAGTGAAGAGAG 480  
 QY 481 GCTTCAGCGCCGAGAGTATCCCATGTTCACCGCCCTGAGAGAGAGCGCCAC 540  
 DB 481 GCTTCAGCGCCGAGAGTATCCCATGTTCACCGCCCTGAGAGAGAGCGCCAC 540  
 QY 541 GACCTGAACACAGATGTTGAACACCGTGGGCGGCGACAGGCGCGCATGCGAGTGTGAAG 600  
 DB 541 GACCTGAACACAGATGTTGAACACCGTGGGCGGCGACAGGCGCGCATGCGAGTGTGAAG 600  
 QY 601 GACACCATCAACAGAGAGGCGCGAGTGGAGCGCGTGCACCCGTCGACCGCGCGCC 660  
 DB 601 GACACCATCAACAGAGAGGCGCGAGTGGAGCGCGTGCACCCGTCGACCGCGCGCC 660  
 QY 661 ATCGCGCGCGCGCAAGTGGCGAGCGCCGCGCGAGAGATGCGCGGCAACCAAGCACC 720  
 DB 661 ATCGCGCGCGCGCAAGTGGCGAGCGCCGCGCGAGAGATGCGCGGCAACCAAGCACC 720  
 QY 721 CTGACAGAGCAGATGCGCTGATGACCAAGCAACCCCGCATCCCGTGGGCGAGCTTAC 780  
 DB 721 CTGACAGAGCAGATGCGCTGATGACCAAGCAACCCCGCATCCCGTGGGCGAGCTTAC 780  
 QY 781 AAGCGGTGATCATCTGGGCTGGAACAAGATGTCGGATGTACGCCCTGAGCATC 840  
 DB 781 AAGCGGTGATCATCTGGGCTGGAACAAGATGTCGGATGTACGCCCTGAGCATC 840

DB 781 AAGCGGTGATCATCTGGGCTGGAACAAGATGTCGGATGTACGCCCTGAGCATC 840  
 QY 841 CTGACATCAACAGAGAGGCGCCCAAGAGGCGCTTTCGCGCATACGTGAGCGCTTTCAG 900  
 DB 841 CTGACATCAACAGAGAGGCGCCCAAGAGGCGCTTTCGCGCATACGTGAGCGCTTTCAG 900  
 QY 901 ACCCTGGCGCGCGAGACAGACCCCAAGAGAGTGAAGACTGATACCGACACCTTGTG 960  
 DB 901 ACCCTGGCGCGCGAGACAGACCCCAAGAGAGTGAAGACTGATACCGACACCTTGTG 960  
 QY 961 GTGCAAGAGCGCAACCCCGACTGCAAGACCACTCTGCGCGCTGTGGCGCGCGCAGC 1020  
 DB 961 GTGCAAGAGCGCAACCCCGACTGCAAGACCACTCTGCGCGCTGTGGCGCGCGCAGC 1020  
 QY 1021 CTGGAGAGATGATGACCGCTTCCAGAGGCGTGGCGCGCCCGACCAAGGCGCGGTG 1080  
 DB 1021 CTGGAGAGATGATGACCGCTTCCAGAGGCGTGGCGCGCCCGACCAAGGCGCGGTG 1080  
 QY 1081 CTGGCGAGAGGATGAGCGAGCGCAACCAAGCGGTGATGATGACAGAGCAACTTCAAG 1140  
 DB 1081 CTGGCGAGAGGATGAGCGAGCGCAACCAAGCGGTGATGATGACAGAGCAACTTCAAG 1140  
 QY 1141 GCGCGCGCGCGCATGCTCAAGTGTCTCACTGCGGCAAGAGGCGCACATGCGCGCAAC 1200  
 DB 1141 GCGCGCGCGCGCATGCTCAAGTGTCTCACTGCGGCAAGAGGCGCACATGCGCGCAAC 1200  
 QY 1201 TGCGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGTGCGCAAGAGGCGCACATGTAAG 1260  
 DB 1201 TGCGCGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGTGCGCAAGAGGCGCACATGTAAG 1260  
 QY 1261 GACTGACCGAGCGCGCAGGCGCAACTCTGGGCAAGACTGAGCGCGACCAAGGCGCGC 1320  
 DB 1261 GACTGACCGAGCGCGCAGGCGCAACTCTGGGCAAGACTGAGCGCGACCAAGGCGCGC 1320  
 QY 1321 CCGCGCAACTTCTGTGACAGCGCGCGCGACCGCGCGCGCGCGCGAGAGCTTCCGC 1380  
 DB 1321 CCGCGCAACTTCTGTGACAGCGCGCGCGCGACCGCGCGCGCGCGCGAGAGCTTCCGC 1380  
 QY 1381 TTGAGAGAGACCAACCCCGCGCAAGAGAGAGAGCAAGAGACCTTGAACCGC 1440  
 DB 1381 TTGAGAGAGACCAACCCCGCGCAAGAGAGAGAGCAAGAGACCTTGAACCGC 1440  
 QY 1441 CTGAAGAGCTGTGTCGAGAGAGACCCCTGAGCGCAGTAA 1479  
 DB 1441 CTGAAGAGCTGTGTCGAGAGAGACCCCTGAGCGCAGTAA 1479

RESULT 2  
 AAL44548  
 ID AAL44548 standard; DNA; 1479 BP.  
 XX  
 AC AAL44548;  
 XX  
 DT 29-AUG-2003 (revised)  
 DT 08-NOV-2002 (first entry)  
 XX  
 DE HIV-1 p55gag polypeptide coding sequence 1.  
 XX  
 XX HIV; de; vaccine; gene; immune response; microparticle;  
 KW adsorbent surface; poly(alpha-hydroxy acid); poly(hydroxy butyric acid);  
 KW polycaprolactone; poly(orthoester); polycyanacrylate; detergent;  
 KW submicron emulsion; viral infection; bacterial infection;  
 KW parasitic infection; HIV-1 p55gag polypeptide.  
 OS Human immunodeficiency virus 1.  
 PN WO20026209-A2.  
 PD 04-APR-2002.  
 XX  
 PF 28-SEP-2001; 2001WO-US030540.  
 XX  
 PR 28-SEP-2000; 2000US-0236105P.

PR		30-AUG-2001; 2001US-0315905P.
XX		(CHIR ) CHIRON CORP.
PA		O'hagan D, O'teen G, Donnelly JJ, Polo JM, Barnett S, Singh M,
PI		Ulmer J, Dubensky TW;
PX		WPI; 2002-519084/55.
DR		A microparticle to which a biologically active macromolecule is adsorbed,
XX		for use as a vaccine composition to treat viral, bacterial or parasitic
PT		infections, comprises a polymer microparticle, a detergent and a
PT		submicron emulsion.
XX		Claim 72; Fig 1; 100pp; English.
PS		The invention relates to a method of raising an immune response in a host
CC		animal. The method of the invention comprises administering a
CC		microparticle that has an adsorbent surface to which a first biologically
CC		active macromolecule (e.g. a nucleic acid) has been adsorbed. The
CC		microparticle comprises a polymer microparticle of poly(alpha-hydroxy
CC		acid), a poly(hydroxy butyric acid), a polycaprolactone, a polyorthoester,
CC		a polyvinylacetate, a detergent, and submicron emulsion. The method/
CC		microparticle of the invention is useful for immunising a host animal
CC		against viral, bacterial or parasitic infections. The present DNA
CC		sequence encodes a HIV-1 p55ag polypeptide. (Updated on 29-AUG-2003 to
CC		standardise OS field)
XX		Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;
SQ		
Query Match	100.0%; Score 1479; DB 6; Length 1479;	
Best Local Similarity	100.0%; Pred. No. 3.2e-185;	
Matches 1479; Conservative	0; Mismatches 0; Indels 0; Gaps 0	
QY	1 ATGGGCGCGCGGCCAGCATCTGTGGCGGGCGCAAGCTGGACGCTGGGAGGCGCATCCGC	60
DB	1 ATGGGCGCGCGGCCAGCATCTGTGGCGGGCGCAAGCTGGACGCTGGGAGGCGCATCCGC	60
QY	61 CTGCGCGCGCGGGGGAAGAAGTGTACTAATGATGAAGCACTTGTTGTGGCCAGCCGCGAG	120
DB	61 CTGCGCGCGCGGGGGAAGAAGTGTACTAATGATGAAGCACTTGTTGTGGCCAGCCGCGAG	120
QY	121 CTGGAGAAGTTGCGCCTGAAACCCCGGCTGTGGAGACCAGCGAGGCTGCAAGCAGATC	180
DB	121 CTGGAGAAGTTGCGCCTGAAACCCCGGCTGTGGAGACCAGCGAGGCTGCAAGCAGATC	180
QY	181 ATCGGCCAGCTGCACACCCCGCTTGAGAACCCGACAGAGAGACTGAAGCCTGTTCAAC	240
DB	181 ATCGGCCAGCTGCACACCCCGCTTGAGAACCCGACAGAGAGACTGAAGCCTGTTCAAC	240
QY	241 ACCGTGGCACCTCTGTACTGTGTCAGCAGAGATCGAGGTCCGCAACAAGAGAGGCC	300
DB	241 ACCGTGGCACCTCTGTACTGTGTCAGCAGAGATCGAGGTCCGCAACAAGAGAGGCC	300
QY	301 CTGGACAAGATCGAGAGAGAGAGACAAGATGCCAGCAGAGAATCCAGCAGGCCCGAGGCC	360
DB	301 CTGGACAAGATCGAGAGAGAGAGACAAGATGCCAGCAGAGAATCCAGCAGGCCCGAGGCC	360
QY	361 GCCGACAAGGGGCAAGGTAGACCAAATTACCCCATGTGTGCAAACTGTGAGGGCCAAGATG	420
DB	361 GCCGACAAGGGGCAAGGTAGACCAAATTACCCCATGTGTGCAAACTGTGAGGGCCAAGATG	420
QY	421 GTGCACAGGCGCATCAGCCCCCGCACCCCTGAACGCTGGGTGTAAGGTGATCGAGGAGAG	480
DB	421 GTGCACAGGCGCATCAGCCCCCGCACCCCTGAACGCTGGGTGTAAGGTGATCGAGGAGAG	480
QY	481 GCCTTCAGCGCCCGAGGTGATCCCATGTTTACCGCCTTAGAGCGAGGCGCCACCCCCGAG	540
DB	481 GCCTTCAGCGCCCGAGGTGATCCCATGTTTACCGCCTTAGAGCGAGGCGCCACCCCCGAG	540
QY	541 GACCTGAACGATGTTGAACACGTTGGGGCGGCACACAGGCCCGCATGCAATGCTGAAG	600
DB	541 GACCTGAACGATGTTGAACACGTTGGGGCGGCACACAGGCCCGCATGCAATGCTGAAG	600

	Accession	Gene	Protein	Length (aa)
QY	601	ABACNCACATCAACGAGAGACCGCCGAGTGGGAAACCGCGTGGACAACCCCGTGCACAGCCGCGCCCC	660	
Db	601	GACACACATCAACGAGAGACCGCCGAGTGGGAAACCGCGTGGACAACCCCGTGCACAGCCGCGCCCC	660	
QY	661	ATCGCCCCCGGACAGATCGCGAGCGCCCGCGGACGCAATCGCGCGGACCAACGACGAC	720	
Db	661	ATCGCCCCCGGACAGATCGCGAGCGCCCGCGGACGCAATCGCGCGGACCAACGACGAC	720	
QY	721	CTGCAGAGACAGATCGCTTGGATGACAGACAACCCCCCATCCCGTGGGCGACATCTAC	780	
Db	721	CTGCAGAGAGCAGATCGCCTTGGATGACAGACAACCCCCCATCCCGTGGGCGACATCTAC	780	
QY	781	AAGGGGTGGATCATCTGGGGGCTGAAACAAGATCGTGGGGATGTACAGCCCGGTGAGATC	840	
Db	781	AAGCGGTGAATCATCTTGGGGCTTGAACAAGATCGTGGGGATGTACAGCCCGGTGAGATC	840	
QY	841	CTGCACATCAACGAGGCGCCCAAGAGAGCCTTCCGCGACTACGTGGACCGCTTCTTCAAG	900	
Db	841	CTGCACATCAACGAGGCGCCCAAGAGAGCCTTCCGCGACTACGTGGACCGCTTCTTCAAG	900	
QY	901	ACCTGCGCGCCGAGCAGAGCAACCAAGAGGTGAAGAACTGSGATGACCGACCTTGTG	960	
Db	901	ACCTGCGCGCCGAGCAGAGCAACCAAGAGGTGAAGAACTGSGATGACCGACCTTGTG	960	
QY	961	GTGCAAGAACGCAACCCCGGACTGGAAGCAATCTCGGGCGCTCTCGGCGCCGGGCGGACG	1020	
Db	961	GTGCAAGAACGCAACCCCGGACTGGAAGCAATCTCGGGCGCTCTCGGCGCCGGGCGGACG	1020	
QY	1021	CTGCAGAGATGATGACCGCCTGCGACGAGCGGTGGGCGGCGCCACGCAAGAGCCGCGCTG	1080	
Db	1021	CTGCAGAGAGATGATGACCGCCTGCGACGAGCGGTGGGCGGCGCCACGCAAGAGCCGCGCTG	1080	
QY	1081	CTGGCCGAGGCGGATGAGCCAGGCCAACAACAGCGTATGATGCAGAAAGCACTTCAAG	1140	
Db	1081	CTGGCCGAGGCGGATGAGCCAGGCCAACAACAGCGTATGATGCAGAAAGCACTTCAAG	1140	
QY	1141	GGCCCCGGGCGCATCGTCAAGTGTCTTCAACTGCGGCGAAGAGAGGACCAATGCGCCGCAAC	1200	
Db	1141	GGCCCCGGGCGCATCGTCAAGTGTCTTCAACTGCGGCGAAGAGAGGACCAATGCGCCGCAAC	1200	
QY	1201	TGCGCGCGCCCCCGCAAGAAAGGCTGTGAAAGTGCGGCAAGAGAGGCCACACAGATGAG	1260	
Db	1201	TGCGCGCGCCCCCGCAAGAAAGGCTGTGAAAGTGCGGCAAGAGAGGCCACACAGATGAG	1260	
QY	1261	GACTGCAACCGAGCGCCAGGCCCACTTCTCTGGGCAAGATCTGGGCCAAGACTGAGCCCAACCA	1320	
Db	1261	GACTGCAACCGAGCGCCAGGCCCACTTCTCTGGGCAAGATCTGAGCCCAACCA	1320	
QY	1321	CCCGGCAACTTCTGCGAGAGCGCGCCGAGGCCAAGCGGCCCGCCCGCGCGAGAGCTTCCGC	1380	
Db	1321	CCCGGCAACTTCTGCGAGAGCGCGCCGAGGCCAAGCGGCCCGCCCGCGCGAGAGCTTCCGC	1380	
QY	1381	TTTCGAGAGACACACCCCGGCGCAAGAGAGAGAGACGCGAGACCTTGAACGAC	1440	
Db	1381	TTTCGAGAGAGACACACCCCGGCGCAAGAGAGAGAGACGCGAGACCTTGAACGAC	1440	
QY	1441	CTGAAGAGCTGTTCGGCAACGACCCCTGTAGCCAGTAA 1479		
Db	1441	CTGAAGAGCTGTTCGGCAACGACCCCTGTAGCCAGTAA 1479		
RESULT 3				
ABL39953				
ID	ABL39953	standard; DNA; 1479 BP.		
XX	ABL39953;			
XX	AC			
XX	DT			
XX	15-MAY-2002	(first entry)		
DE	Synthetic Gag polynucleotide sequence SBQ ID NO:3.			
XX	Human immunodeficiency virus type C; antigenic HIV type C protein;			





```
RESULT 4
ADM73758
ID ADM73758 standard; DNA; 1479 BP.
XX
AC ADM73758;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #1.
XX
XX HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
XX HIV type C protein; immunostimulant.
XX
OS Human immunodeficiency virus 1.
XX
PN US2003223961-A1.
XX
PD 04-DEC-2003.
XX
PE 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGE/) MEGEDE J Z.
PA (BARN/) BARNETT S W.
PA (ENGE/) ENGELBRECHT S.
PA (RENS/) RENSBURG B J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg BJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV pol polypeptide, useful in eliciting an immune response, in DNA
PT immunisation, generating of packaging cell lines or in producing HIV type
PT C proteins.
XX
PS Example 1; SEQ ID NO 3; 160pp; English.
XX
XX The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX
SQ Sequence 1479 BP; 325 A; 529 C; 463 G; 162 T; 0 U; 0 Other;
Query Match 100.0%; Score 1479; DB 12; Length 1479;
Best Local Similarity 100.0%; Pred. No. 3,2e-185;
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGGCGCCGCGCGCATCTGCGCGCGCGCGCAAGCTGACGACCTGGAGGCGCATCCGC 60
DB 1 ATGGGCGCCGCGCGCATCTGCGCGCGCGCGCAAGCTGACGACCTGGAGGCGCATCCGC 60
QY 61 CTGCGCGCCGCGCGCAAGAGTGCTACATGATGAAACACCTGTGTGTGGCCAGCCGCGAG 120
DB 61 CTGCGCGCCGCGCGCAAGAGTGCTACATGATGAAACACCTGTGTGTGGCCAGCCGCGAG 120
QY 121 CTGGAAGAAATTTGGCTTGAACCCCGGCTGTGTGGAACAACAAGGAGGCTGGCAAGCATC 180
DB 121 CTGGAAGAAATTTGGCTTGAACCCCGGCTGTGTGGAACAACAAGGAGGCTGGCAAGCATC 180
QY 181 ATCCGCGACGCTGACCCCGCTGTGAGACGCGGAGGAGGAGCTGAAGAGCTGTTCAC 240
DB 181 ATCCGCGACGCTGACCCCGCTGTGAGACGCGGAGGAGGAGCTGAAGAGCTGTTCAC 240
```

```
DB 181 ATCCGCGACGCTGACCCCGCTGTGAGACGCGGAGGAGGAGCTGAAGAGCTGTTCAC 240
QY 241 ACCGTGGCCACCTCTGTACTGCGTGCACGAGAAAGATCGAGTCTCCGCAACCAAGAGGCC 300
DB 241 ACCGTGGCCACCTCTGTACTGCGTGCACGAGAAAGATCGAGTCTCCGCAACCAAGAGGCC 300
QY 301 CTGGAAGAAATGAGAGGAGGAGCAAGATGCGCAGCAAGATGATCAGAGGCCGAGGCC 360
DB 301 CTGGAAGAAATGAGAGGAGGAGCAAGATGCGCAGCAAGATGATCAGAGGCCGAGGCC 360
QY 361 GCCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
DB 361 GCCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
QY 421 GTGCAACAGGCGATGAGGCGCCGCGACCTGTGAACGCGCTGGGTGAAGTGTGAGAGAG 480
DB 421 GTGCAACAGGCGATGAGGCGCCGCGACCTGTGAACGCGCTGGGTGAAGTGTGAGAGAG 480
QY 481 GCCTTCAAGCGCGGAGGTGATCCCATGTTCAACCGGCTGAGGAGGAGGAGGAGGAGGAG 540
DB 481 GCCTTCAAGCGCGGAGGTGATCCCATGTTCAACCGGCTGAGGAGGAGGAGGAGGAGGAG 540
QY 541 GACCTGAACACGATGTTGAACAACGCTGGGCGGCAACAGGCCCGCATGAGATGCTGAAG 600
DB 541 GACCTGAACACGATGTTGAACAACGCTGGGCGGCAACAGGCCCGCATGAGATGCTGAAG 600
QY 601 GACACCATCAACAGAGAGGCGCGCGAGTGGAGACCGCGTGCACCCCGTGCACCCCGGCC 660
DB 601 GACACCATCAACAGAGAGGCGCGCGAGTGGAGACCGCGTGCACCCCGTGCACCCCGGCC 660
QY 661 ATGCGCCCGCGCGAGATGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
DB 661 ATGCGCCCGCGCGAGATGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 721 CTGCAAGAGCAGATGCGCTGATGAACAACACCCCGCATCCCGTGGGCGCATCTAC 780
DB 721 CTGCAAGAGCAGATGCGCTGATGAACAACACCCCGCATCCCGTGGGCGCATCTAC 780
QY 781 AAGCGGTGATATCTGTGGCTTGAACAAGATCGTGGGATGATACAGCCCGTGAACATC 840
DB 781 AAGCGGTGATATCTGTGGCTTGAACAAGATCGTGGGATGATACAGCCCGTGAACATC 840
QY 841 CTGGAATCAACAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
DB 841 CTGGAATCAACAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
QY 901 ACCCTGCGCGCGCGAGCAGAGCAACCGAGAGTGAAGAACTGGATGACGACACCTGCTG 960
DB 901 ACCCTGCGCGCGCGAGCAGAGCAACCGAGAGTGAAGAACTGGATGACGACACCTGCTG 960
QY 961 GTGCAAGACGCAACCCCGCATCTGCAAGACATCTGCGGCTGTGGCCCGCGCGCGCG 1020
DB 961 GTGCAAGACGCAACCCCGCATCTGCAAGACATCTGCGGCTGTGGCCCGCGCGCGCG 1020
QY 1021 CTGGAAGAGATGATACCGGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
DB 1021 CTGGAAGAGATGATACCGGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
QY 1081 CTGGAAGAGATGATACCGGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
DB 1081 CTGGAAGAGATGATACCGGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
QY 1141 GCGCGCGCGCGCGCATGTCTCAAGTCTTCAACTGCGGCAAGAGGCGCAATCGCCGCAAC 1200
DB 1141 GCGCGCGCGCGCGCATGTCTCAAGTCTTCAACTGCGGCAAGAGGCGCAATCGCCGCAAC 1200
QY 1201 TGGCGGCGCGCGCGCAAGAGGCGCTGGAAGTGGCGCAAGAGGAGGAGGAGGAGGAG 1260
DB 1201 TGGCGGCGCGCGCGCAAGAGGCGCTGGAAGTGGCGCAAGAGGAGGAGGAGGAGGAG 1260
QY 1261 GACTGCAACGAGCGCGCGCAACTTCTGTGGCAAGATCTGGCCGACCAAGAGGCGCG 1320
DB 1261 GACTGCAACGAGCGCGCGCAACTTCTGTGGCAAGATCTGGCCGACCAAGAGGCGCG 1320
```





```
QY 961 GTGCAGAAAGCCGCAACCCGAGCTGCAAGACATCTGCGCGCTCTGAGCCCGCGGCCAGC 1020
    |||||
Db 967 GTGCAGAAAGCCGCAACCCGAGCTGCAAGACATCTGCGCGCTCTGAGCCCGCGGCCAGC 1026
QY 1021 CTGAGAGAGATGATGATACCGGCTGCGAGGGCGCTGGGGCGGCCCAAGCAAGGCCCGGCTG 1080
    |||||
Db 1027 CTGAGAGAGATGATGATACCGGCTGCGAGGGCGCTGGGGCGGCCCAAGCAAGGCCCGGCTG 1086
QY 1081 CTGAGCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
    |||||
Db 1087 CTGAGCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1146
QY 1141 GGGCGCGCGCGCGATGCTCAAGTGTCTTCAACTGCGCGCAAGAGAGGCCCACTGCGCGCAC 1200
    |||||
Db 1147 GGGCGCGCGCGCGATGCTCAAGTGTCTTCAACTGCGCGCAAGAGAGGCCCACTGCGCGCAC 1206
QY 1201 TGGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
    |||||
Db 1207 TGGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
QY 1261 GACTGACCGAGAGCGCGAGCGCAACTTCTGCGCAAGATCTGCGCGCAAGCGCAAGAGAGCGCGC 1320
    |||||
Db 1267 GACTGACCGAGAGCGCGAGCGCAACTTCTGCGCAAGATCTGCGCGCAAGCGCAAGAGAGCGCGC 1326
QY 1321 CCGCGCACTTCTCTGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
    |||||
Db 1327 CCGCGCACTTCTCTGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386
QY 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
    |||||
Db 1387 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
QY 1441 CTGAGAGAGCGCTGTTGGCAACGACCCCTGAGCGCACTAA 1479
    |||||
Db 1447 CTGAGAGAGCGCTGTTGGCAACGACCCCTGAGCGCACTAA 1485

RESULT 7
ID ACA03523 standard; DNA; 4419 BP.
XX
AC ACA03523;
XX
DE 22-MAY-2003 (first entry)
XX
DE Synthetic DNA encoding immunogenic HIV peptide #6.
XX
KM Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
XX cellular immune response; gene delivery vector; DNA immunisation; de.
XX
XX Synthetic.
XX
XX OS
XX PN WO2003004657-A1.
XX
XX PD 16-JAN-2003.
XX
XX PF 05-JUL-2002; 2002WO-US021421.
XX
XX PR 05-JUL-2001; 2001US-0303192P.
XX 31-AUG-2001; 2001US-0316860P.
XX 16-JAN-2002; 2002US-0349728P.
XX 16-JAN-2002; 2002US-0349733P.
XX 16-JAN-2002; 2002US-0349871P.
XX
XX PA (CHIR ) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Llan Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
```

PT polypeptides, useful as immunogenic compositions or vaccines for  
PT generating humoral or cellular immune responses against HIV in a subject,  
XX especially humans.

Example 1; Fig 11; 262bp; English.

XX The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation, generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response) in, a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide

XX Sequence 4419 BP; 980 A; 1582 C; 1371 G; 486 T; 0 U; 0 Other;

Query Match 99.9%; Score 1477.4; DB 8; Length 4419;  
Best Local Similarity 99.9%; Pred. No. 4.3e-185;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGGCGCGCGCGCGAGCATCTGCGCGCGAGAGCTGAGCGCTGAGAGCGCATTCGC 60
    |||||
Db 7 ATGGGCGCGCGCGCGAGCATCTGCGCGCGAGAGCTGAGCGCGCATTCGC 66
QY 61 CTGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
    |||||
Db 67 CTGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126
QY 121 CTGAGAGAGTTGCGCTGAAACCCCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
    |||||
Db 127 CTGAGAGAGTTGCGCTGAAACCCCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186
QY 181 ATCCGCGAGCTGCAACCCCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
    |||||
Db 187 ATCCGCGAGCTGCAACCCCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
QY 241 ACCGTGGCACCTCTGTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
    |||||
Db 247 ACCGTGGCACCTCTGTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 306
QY 301 CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
    |||||
Db 307 CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366
QY 361 GCCGCAAGAGGCAAGGTGAGCGCAAGACTTACCTTCTGTGCAAGAGAGAGAGAGAGAGAGAGAG 420
    |||||
Db 367 GCCGCAAGAGGCAAGGTGAGCGCAAGACTTACCTTCTGTGCAAGAGAGAGAGAGAGAGAGAGAG 426
QY 421 GTGCAACAGGCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
    |||||
Db 427 GTGCAACAGGCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486
QY 481 GCGTTCAAGCGCGAGGTGATCCCATGTTTCAACCGCGCTGAGCGAGAGAGCGCGCGCGCGCGAG 540
    |||||
Db 487 GCGTTCAAGCGCGAGGTGATCCCATGTTTCAACCGCGCTGAGCGAGAGAGCGCGCGCGCGCGAG 546
QY 541 GACTGAAACAGATTTGAAACAGCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
    |||||
Db 547 GACTGAAACAGATTTGAAACAGCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606
QY 601 GACACCATCAACAGAGAGAGCGCGCGAGTGGAGACCGCGTGCACCGCGTGCACCGCGCGCGCGCG 660
    |||||
Db 607 GACACCATCAACAGAGAGAGCGCGCGAGTGGAGACCGCGTGCACCGCGTGCACCGCGCGCGCGCG 666
QY 661 ATGCGCGCGCGCGAGATGCGAGCGCGCGCGCGCGAGCATGCGCGGCAACCAAGCAAC 720
    |||||
Db 667 ATGCGCGCGCGCGAGATGCGAGCGCGCGCGCGCGAGCATGCGCGGCAACCAAGCAAC 726
```

```

QY 721 CTGACGAGACGATCGCTGATGACCAAGACCCCGCCCGGCGACATCTAC 780
DB 727 CTGACGAGACGATCGCTGATGACCAAGACCCCGCCCGGCGACATCTAC 786
QY 781 AAGCGGTGATCATCTGCGGCTGAAACAAGTGTGCGGATGTACAGCCCGGTAGATC 840
DB 787 AAGCGGTGATCATCTGCGGCTGAAACAAGTGTGCGGATGTACAGCCCGGTAGATC 846
QY 841 CTGACATCAAGACGAGGCCCCCAAGAGCCCTTCCGAGTACTAGTGAACCGCTTTCAAG 900
DB 847 CTGACATCAAGACGAGGCCCCCAAGAGCCCTTCCGAGTACTAGTGAACCGCTTTCAAG 906
QY 901 ACCCTGCGGCGGAGACGACACCCAGAGGTGAAGAATGTAGTACCGACCCCTGTG 960
DB 907 ACCCTGCGGCGGAGACGACACCCAGAGGTGAAGAATGTAGTACCGACCCCTGTG 966
QY 961 GTGCAAAAGCCAAACCCCGACTGCAAGACATCTGCGGCTTCTCGGCCCCGCGCAGC 1020
DB 967 GTGCAAAAGCCAAACCCCGACTGCAAGACATCTGCGGCTTCTCGGCCCCGCGCAGC 1026
QY 1021 CTGAGAGAGATGATGACCCCTGTCAGAGGCGCTGCGGCGGCGGCGGCGGCGGCGG 1080
DB 1027 CTGAGAGAGATGATGACCCCTGTCAGAGGCGCTGCGGCGGCGGCGGCGGCGGCGG 1086
QY 1081 CTGAGCGAGCGATGAGCCAGGCGCAACACAGCGTGTATGAGAGAGCAACTTCAAG 1140
DB 1087 CTGAGCGAGCGATGAGCCAGGCGCAACACAGCGTGTATGAGAGAGCAACTTCAAG 1146
QY 1141 GCGCCCGCGGCGATCGTCAAGTCTTCAACTGCGGCGAGAGGCGCAATCGCCGCGAC 1200
DB 1147 GCGCCCGCGGCGATCGTCAAGTCTTCAACTGCGGCGAGAGGCGCAATCGCCGCGAC 1206
QY 1201 TCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
DB 1207 TCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1266
QY 1261 GACTGACACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
DB 1267 GACTGACACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1326
QY 1321 CCGCGCAATCTCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
DB 1327 CCGCGCAATCTCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1386
QY 1381 TTGAGAGAGACCAACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
DB 1387 TTGAGAGAGACCAACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1446
QY 1441 CTGAAGAGCTGTGGGCAAGACCCCGTGAAGCACTAA 1479
DB 1447 CTGAAGAGCTGTGGGCAAGACCCCGTGAAGCACTAA 1485

```

```

RESULT 8
ADCI3240
ID ADCI3240 strand; DNA; 4419 BP.
XX
XX ADCI3240;
XX
XX 18-DEC-2003 (first entry)
XX
XX DNA of HIV construct GagRTmutTatRevNef_C SEQ ID NO 19.
XX
XX expression cassette; HIV Gag; Env; Int; Nef; p15RaseH; Pol; Tat; Prot;
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
XX Human immunodeficiency virus.
XX
XX WO2003004620-A2.
XX
XX 16-JAN-2003.
XX
XX 05-JUL-2002; 2002WO-US021420.
XX
XX PF

```

```

XX 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
XX (CHIR ) CHIRON CORP.
XX PA (UYST-) UNIV STELLENBOSCH.
XX
XX Zur Megele J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ,
XX WPI; 2003-221593/21.
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RaseH, Pol, Tat,
XX Prot, or Rev polypeptide, useful for immunization, or generating
XX packaging cell lines.
XX
XX Disclosure; Fig 16; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX expression cassette can be used to treat HIV type C by gene therapy or
XX used in the development of a vaccine. The gene delivery vector is
XX administered intramuscularly, intravenously, intranasally,
XX subcutaneously, intradermally, transdermally, intravaginally,
XX intrarectally, orally or intravenously. The expression cassette is useful
XX for immunisation, generating packaging cell lines and producing HIV
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX Type C related sequence of the invention.
XX
XX Sequence 4419 BP; 980 A; 1582 C; 1371 G; 486 T; 0 U; 0 Other;

```

```

Query Match 99.9%; Score 1477.4; DB 10; Length 4419;
Best Local Similarity 99.9%; Pred. No. 4.3e-185;
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 ATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
DB 7 ATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 66
QY 61 CTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 67 CTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 126
QY 121 CTGAAGAGTTGCGCTGAAACCCCGGCTGCTGAGACCAAGGAGGCTGCAAGATC 180
DB 127 CTGAAGAGTTGCGCTGAAACCCCGGCTGCTGAGACCAAGGAGGCTGCAAGATC 186
QY 181 ATCCGCGAGCTGCAACCCCGGCTGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 187 ATCCGCGAGCTGCAACCCCGGCTGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 246
QY 241 ACCGTGCGCACCTGTACTGCTGACGAGAAAGATCGAGTCCGCGACCAAGAGGCGC 300
DB 247 ACCGTGCGCACCTGTACTGCTGACGAGAAAGATCGAGTCCGCGACCAAGAGGCGC 306
QY 301 CTGAACAAGTGAAGAGAGCAAAAGTCCGACGAGAGATCAAGAGGCGCGAGGCC 360
DB 307 CTGAACAAGTGAAGAGAGCAAAAGTCCGACGAGAGATCAAGAGGCGCGAGGCC 366
QY 361 GCGGCAAGGAGGAGAGTGAAGCAATACCCCATCTGCGAGAACTGAGGCGGCGAGT 420
DB 367 GCGGCAAGGAGGAGAGTGAAGCAATACCCCATCTGCGAGAACTGAGGCGGCGAGT 426
QY 421 GTGACACAGGCGCATGAGCCCGCGACCTGAAAGCGCTGAGTGAAGTATCGAGAGAG 480
DB 427 GTGACACAGGCGCATGAGCCCGCGACCTGAAAGCGCTGAGTGAAGTATCGAGAGAG 486
QY 481 GCTTACGCGCGGAGGTATCCCATGTTCAACCGGCTGAGGAGGCGGCGGCGGCGG 540
DB 487 GCTTACGCGCGGAGGTATCCCATGTTCAACCGGCTGAGGAGGCGGCGGCGGCGG 546

```

541 GACCTGAACAGCATGTTGAACACCGTGGGCGGCGACACAGGCCGCAATGAGTGTGAAG 600  
547 GACCTGAACAGCATGTTGAACACCGTGGGCGGCGACACAGGCCGCAATGAGTGTGAAG 606  
601 GACACATCAACAGAGAGGCGCGGAGTGGAGCCGCGTGCACCCCGTGCACCGCGCGCC 660  
607 GACACATCAACAGAGAGGCGCGGAGTGGAGCCGCGTGCACCCCGTGCACCGCGCGCC 666  
661 ATCGCCCCCGGAGATGCGAGAGCGCGCGGAGAGATGAGCGGCGGACCAACAGCAC 720  
667 ATCGCCCCCGGAGATGCGAGAGCGCGGAGAGATGAGCGGCGGACCAACAGCAC 726  
721 CTGCAAGAGACATGCGCTGATGACAGCAACCCCGCATCCCGTGGGCGACATCTAC 780  
727 CTGCAAGAGACATGCGCTGATGACAGCAACCCCGCATCCCGTGGGCGACATCTAC 786  
781 AAGCGGTGATCATCTGGGCTGAAACAAGATGTCGGATGACAGCCCGTGAAGCATC 840  
787 AAGCGGTGATCATCTGGGCTGAAACAAGATGTCGGATGACAGCCCGTGAAGCATC 846  
841 CTGCAATCAACAGAGGCGCGGAGAGCGCGGAGATGAGTGTGAAGTGTGATGAGTGTG 900  
847 CTGCAATCAACAGAGGCGCGGAGAGCGCGGAGATGAGTGTGAAGTGTGATGAGTGTG 906  
901 ACCCTGCGCGCGAGACAGACACCAAGAGTGAAGAACTGGATGACGACACCTGCTG 960  
907 ACCCTGCGCGCGAGACAGACACCAAGAGTGAAGAACTGGATGACGACACCTGCTG 966  
961 GTGCAAAACGCAACCCCGACTGCAAGACATCTGGCGCTCTGGCGCGCGCGCAAC 1020  
967 GTGCAAAACGCAACCCCGACTGCAAGACATCTGGCGCTCTGGCGCGCGCGCAAC 1026  
1021 CTGGAAGAGATGATGACCGGCTGCAAGAGCGGCGGCGCGGCGGCGGCGGCGGCG 1080  
1027 CTGGAAGAGATGATGACCGGCTGCAAGAGCGGCGGCGGCGGCGGCGGCGGCGG 1086  
1081 CTGCGCGAGGCGATGAGCGGCGGCAACACAGCGTGTATGATGACAGAGAGCACTTCA 1140  
1087 CTGCGCGAGGCGATGAGCGGCGGCAACACAGCGTGTATGATGACAGAGAGCACTTCA 1146  
1141 GCGCGCGCGCGATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1200  
1147 GCGCGCGCGCGATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1206  
1201 TCGCGCGCGCGCGGAGAGAGGCGTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGA 1260  
1207 TCGCGCGCGCGCGGAGAGAGGCGTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGA 1266  
1261 GACTGCAACGAGCGGCGGCAACTTCTGTGGGCAAGATGTGGCCGACCAAGAGGCGG 1320  
1267 GACTGCAACGAGCGGCGGCAACTTCTGTGGGCAAGATGTGGCCGACCAAGAGGCGG 1326  
1321 CCGGCGCACTTCTGTGAAGCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
1327 CCGGCGCACTTCTGTGAAGCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1386  
1381 TTGCGAGAGACACCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
1387 TTGCGAGAGACACCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
1441 CTGAAGAGCTGTGTGCGCAACGAGCGCGTGAAGCACTAA 1479  
1447 CTGAAGAGCTGTGTGCGCAACGAGCGCGTGAAGCACTAA 1485

RESULT 9  
ADCI3257  
ID ADCI3257 standard; DNA; 4615 BP.  
XX  
AC ADCI3257;  
XX  
DT 18-DEC-2003 (first entry)  
XX

DE DNA of HIV construct gp160mod-TV1-dV2-gagmod-BW965 SEQ ID NO 36.  
XX  
XX expression cassette; HIV Gag; Env; Int; Nef; p15RaseH; Pol; Tat; Prot;  
KM Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX  
XX Human immunodeficiency virus.  
OS  
XX  
XX WO200304620-A2.  
PN  
XX  
XX 16-JAN-2003.  
PD  
XX  
XX 05-JUL-2002; 2002WO-US021420.  
PF  
XX  
XX 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
XX (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
XX Zur Megede J, Barnett SM, Lian Y, Engelbrecht S, Van Rensburg EJ;  
PI WPI, 2003-221593/21.  
XX  
XX  
XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX  
XX  
XX Disclosure; Fig 33; 301bp; English.  
PS  
XX  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intravenously, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX  
XX  
SQ Sequence 4615 BP; 1030 A; 1587 C; 1354 G; 644 T; 0 U; 0 Other;  
Query Match 99.9%; Score 1477.4; DB 10; Length 4615;  
Best Local Similarity 99.9%; Pred. No. 4.3e-185;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60  
DB 3137 ATGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3196  
QY 61 CTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
DB 3197 CTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3256  
QY 121 CTGGAAGAGTTCGCTGTAACCCCGGCTGTCGAGACGAGAGGCTGCAAGCATC 180  
DB 3257 CTGGAAGAGTTCGCTGTAACCCCGGCTGTCGAGACGAGAGGCTGCAAGCATC 3316  
QY 181 ATCGCGAGCTGCAACCGCGCGGCTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
DB 3317 ATCGCGAGCTGCAACCGCGCGGCTGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3376  
QY 241 ACCGTGCGCACCTGTACTGCTGCTGCAAGAGATGAGTCCGCGACACCAAGAGGCGC 300  
DB 3377 ACCGTGCGCACCTGTACTGCTGCTGCAAGAGATGAGTCCGCGACACCAAGAGGCGC 3436  
QY 301 CTGCAAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 3437 CTGCAAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3496

OY	361	GGCGACAAGGGGCAAGGTGGCCAGAACTACCCCATCTGACAAACCTGGAGGGCCAGATG	420
Db	3497	GGCGACAAGGGGCAAGGTGGCCAGAACTACCCCATCTGACAAACCTGGAGGGCCAGATG	3556
OY	421	GTGCACAAGGCCATCAGGCCCGCCACCCCTGAACCGCTGGGTGAAGGTGATTCAGAGAAAG	480
Db	3557	GTGCACAAGGCCATCAGGCCCGCCACCCCTGAACCGCTGGGTGAAGGTGATTCAGAGAAAG	3611
OY	481	GCCTTCAGCCCCGAGGTGATTCCTCATGTTCAACCGCTTGAGCGAGGGCGCAACCCCGAG	540
Db	3617	GCCTTCAGCCCCGAGGTGATTCCTCATGTTCAACCGCTTGAGCGAGGGCGCAACCCCGAG	3678
OY	541	GACCTGAACACGATGTGTGAACAACCGTGGGGCGCCACAGGGCCGCAATGCAAGATGTGTAAG	600
Db	3677	GACCTGAACACGATGTGTGAACAACCGTGGGGCGCCACAGGGCCGCAATGCAAGATGTGTAAG	3733
OY	601	GACACCATCAACGAGAGAGCGCCCGCAGTGGAGACCGCGTGCACCCCGTGCACCCCGGCC	660
Db	3737	GACACCATCAACGAGAGAGCGCCCGCAGTGGAGACCGCGTGCACCCCGTGCACCCCGGCC	3798
OY	661	ATCGCCCCCGCCAGATGCGCGAGCCCGCGCAGCGACATGCGCGGCACACAGCAACC	720
Db	3797	ATCGCCCCCGCCAGATGCGCGAGCCCGCGCAGCGACATGCGCGGCACACAGCAACC	3853
OY	721	CTGCAGAGACGAGATGCGCTGTGATGACAGCAACCCCGCATCCCTGGGGCGACATCTAC	780
Db	3857	CTGCAGAGACGAGATGCGCTGTGATGACAGCAACCCCGCATCCCTGGGGCGACATCTAC	3913
OY	781	AAGCGGTGGATCATCTCTGGGCTTGAAACAAGATCGTGGATGTACAGCCCGGTGACATC	840
Db	3917	AAGCGGTGGATCATCTCTGGGCTTGAAACAAGATCGTGGATGTACAGCCCGGTGACATC	3978
OY	841	CTGCACATCAAGCAGGCGCCCAAGAGCCCTTTCGCGCATCAGTGTGACCGCTTCTTCAAG	900
Db	3977	CTGCACATCAAGCAGGCGCCCAAGAGCCCTTTCGCGCATCAGTGTGACCGCTTCTTCAAG	4033
OY	901	ACCTCGCGCGCCGAGACAGACACCCAGAGGTGTGAAGAACTGGATGACCGACACTTGGTG	960
Db	4037	ACCTCGCGCGCCGAGACAGACACCCAGAGGTGTGAAGAACTGGATGACCGACACTTGGTG	4098
OY	961	GTGCAGAACGGCCAAACCCCGCATGCAAGAACATCTCGCGCGCTCTGGGCCCGCGGCGCACG	1020
Db	4097	GTGCAGAACGGCCAAACCCCGCATGCAAGAACATCTCTGGCGCTTCTGGGCCCGCGGCGCACG	4155
OY	1021	CTGCAGAGATGATGACCGCTTGCCAGGCGCTTGCGCGCGCCAGCCACAGAGCCCGCGTG	1080
Db	4157	CTGCAGAGATGATGACCGCTTGCCAGGCGCTTGCGCGCGCCAGCCACAGAGCCCGCGTG	4216
OY	1081	CTGGCCGAGCGGATGACCGAGCCAAACACAGCTGATGATGACAGAAAGCACTTCAAG	1140
Db	4217	CTGGCCGAGCGGATGACCGAGCCAAACACAGCTGATGATGACAGAAAGCACTTCAAG	4276
OY	1141	GGCCCCGGCGGATGTGCAAGTGTCTCAATCGCGGCAAGAGGGGCAATGCGCGCGCAAC	1200
Db	4277	GGCCCCGGCGGATGTGCAAGTGTCTCAATCGCGGCAAGAGGGGCAATGCGCGCGCAAC	4333
OY	1201	TGCGCGCGCCCCCGCAAGAGGGCTGTGGAAGTCCGGCAAGAGGGCCACCAAGATGAAG	1260
Db	4337	TGCGCGCGCCCCCGCAAGAGGGCTGTGGAAGTCCGGCAAGAGGGCCACCAAGATGAAG	4398
OY	1261	GACTGTCAACGAGCGCCAGGGCCAACTTCTGGGCAAGATCTGGGCCAGGCCACAAAGGGCGG	1320
Db	4397	GACTGTCAACGAGCGCCAGGGCCAACTTCTGGGCAAGATCTGGGCCAGGCCACAAAGGGCGG	4455
OY	1321	CCCGGCAACTTCTGTGCAAGCGCGCCCGAGGCCACCGCCCCCGCGGAGGCTTTCGG	1380
Db	4457	CCCGGCAACTTCTGTGCAAGCGCGCCCGAGGCCACCGCCCCCGCGGAGGCTTTCGG	4516
OY	1381	TTTCGAGAGACCAACCCCGCGCCAGAGAGGAGACAAAGACACCGCGAACCCTTGACAGC	1440
Db	4517	TTTCGAGAGACCAACCCCGCGCCAGAGAGGAGACAAAGACACCGCGAACCCTTGACAGC	4578
OY	1441	CTGAAGAGCTGTGTCGCAACGACCCCTTGAGCCAGTAA	1479

Db 4577 CTGAAGAGCCTGTTCCGCAACGACCCCTGAGCCAATTA 4615

RESULT 10  
ADCI3259  
ID ADCI3259 standard; DNA; 4702 BP.

AC ADC13259;  
VY

DT 18-DEC-2003 (first entry)  
 XY

DNA of HIV construct gp160mod-TV1-gagmod-BW965 SEQ ID NO 38.

KM expression cassette; HIV Gag; Env; Int; Nef; p15KasabH; Pol; Tat; Prot; Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

OS Human immunodeficiency virus.

PN WO2003004620-A2.

PD 16-JAN-2003.

PF 05-JUL-2002; 2002WO-US021420.

PR 05-JUL-2001; 2001US-0303192P.

PR 16-JAN-2002; 2002US-0349871P.

PA (CHIR ) CHIRON CORP.  
PA (IVCT - ) INTV STEEL ENROSCU

XX Zuz Maccado T Barnett CW Tinsy Engelbrecht C Van Rensburg ET

XX  
WDT: 2003-221E03/21

XX  
XX  
New experience comes in a variety of different guises and

PT polypeptide including an HIV Gag, Env, Int, Nef, p15RNaseH, Pol, Tat, Rev or Rev polypeptide useful for immunization or generating

PT packaging cell lines.

PS Disclosure; Fig 35; 301pp; English.

CC The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag

CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel

used in the development of a vaccine. The gene delivery vector is

CC subcutaneously, intradermally, transdermally, intravaginally, intracervically, orally or intravenously. The compound can be used

CC for immunisation, generating packaging cell lines and producing HIV polynucleotides. This polynucleotide sequence represents the DNA of an HIV

CC Type C related sequence of the invention.  
 YY

**SQ** Sequence 4702 BP; 1057 A; 1617 C; 1371 G; 657 T; 0 U; 0 Other;

Query Match	99.9%	Score 1477.4;	DB 10;	Length 4702;
Best Local Similarity	99.9%	Pred NO 4	3e-185.	

Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGGCCCCGCGCCAGCATCTGCGCGGGCGGCAAGCTGGACGCCCTGGAGCGCATCCGC 60

Db 3224 ATGGGCGCCCGCAGCATCCTGCGGGCGGCAAGCTGGACGCCCTGGGAGCCATCCGC 3283

61 CTGCGCCCCGGCGCAAGAGTCTACATGATGAAGCACCTGGTGTGGCCAGCCGGAG 120

Db 3284 CTGCCCCCGCCGCAAGAGTGTCTACATGATGAAGCACCTGGTGTGGCCAGCCGCGAG 3343

121 CTGGAGAGTTTCGCCCTGACCCCGGCTCTGGAGACCGGAGGGCTGCACGACGATC 180

D5 334 CTGGAGAAGTTCGCCCTGACCCCCGGCCTGCTGGAGACCAAGCAGGGCTGCACGACATC 3403



```
QY 181 ATCCGACAGTGCACCCCGCCCTGCAAGACCGGACGAGGAGCTGAAGACCTTGTCAAC 240
DB 3404 ATCCGACAGTGCACCCCGCCCTGCAAGACCGGACGAGGAGCTGAAGACCTTGTCAAC 3463
QY 241 ACCGTGGCCACCTGTACTGCGTGCAAGAGAGTGAAGTCCGACACCAAGAGAGCC 300
DB 3464 ACCGTGGCCACCTGTACTGCGTGCAAGAGAGTGAAGTCCGACACCAAGAGAGCC 3523
QY 301 CTGGAAGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 3524 CTGGAAGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3583
QY 361 GCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 3584 GCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3643
QY 421 GTGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 3644 GTGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3703
QY 481 GCCTTCAGCCCGGAGAGTGAATCCCATGTTCACCGCCCTGAAGAGAGAGAGAGAG 540
DB 3704 GCCTTCAGCCCGGAGAGTGAATCCCATGTTCACCGCCCTGAAGAGAGAGAGAGAG 3763
QY 541 GACCTGAACAGATGTTGAACAACCGTGAGGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 3764 GACCTGAACAGATGTTGAACAACCGTGAGGAGAGAGAGAGAGAGAGAGAGAGAG 3823
QY 601 GACATCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 3824 GACATCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3883
QY 661 ATGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 3884 ATGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3943
QY 721 CTGCAAGAGAGAGATGCTGATGAACAGCAACCCGCCCATCCCGTGAGAGATCTAC 780
DB 3944 CTGCAAGAGAGAGATGCTGATGAACAGCAACCCGCCCATCCCGTGAGAGATCTAC 4003
QY 781 AAGCGGTGATCATCTCTGGGCTTGAACAAGATGTCGGATTTACAGCCCGCTGACATC 840
DB 4004 AAGCGGTGATCATCTCTGGGCTTGAACAAGATGTCGGATTTACAGCCCGCTGACATC 4063
QY 841 CTGCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 4064 CTGCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4123
QY 901 ACCCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 4124 ACCCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4183
QY 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 4184 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4243
QY 1021 CTGGAAGAGATGATACCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 4244 CTGGAAGAGATGATACCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4303
QY 1081 CTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 4304 CTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4363
QY 1141 GAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 4364 GAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4423
QY 1201 TGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 4424 TGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4483
QY 1261 GACTGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
```

```
DB 4484 GACTGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4543
QY 1321 CCCGCAACTTCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 4544 CCCGCAACTTCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4603
QY 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 4604 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4663
QY 1441 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479
DB 4664 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4702
```

## RESULT 11

ACN03522

ID ACA03522 standard; DNA; 4716 BP.

XX ACA03522;

DT 22-MAY-2003 (first entry)

XX Synthetic DNA encoding immunogenic HIV peptide #5.

KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;

KW gene therapy; packaging cell line; humoral immune response;

KW cellular immune response; gene delivery vector; DNA immunisation; ds.

OS Synthetic.

XX WO200304657-A1.

XX 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021421.

XX 05-JUL-2001; 2001US-0303192P.

XX 31-AUG-2001; 2001US-0316860P.

XX 16-JAN-2002; 2002US-0349728P.

XX 16-JAN-2002; 2002US-0349733P.

XX 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

XX Zur Megede J, Barnett SW, Lian Y;

XX WPI; 2003-221602/21.

XX Example 1; Fig 10; 262pp; English.

The invention describes a synthetic polynucleotide encoding 2 or more immunogenic HIV polypeptides, where at least 2 of the polypeptides are derived from different HIV subtypes. The polynucleotide is useful for immunisation, generation of packaging cell lines, or production of HIV polypeptides. The polynucleotide and its encoded proteins are useful as immunogenic compositions or vaccines for generating humoral or cellular immune responses against HIV in a subject, or for inducing neutralising antibodies against HIV. The gene delivery vector comprising the polynucleotide is also useful for DNA immunisation of, or for generating an immune response (e.g. a humoral or cellular immune response) in, a subject such as a mammal, particularly a human. This sequence encodes a human immunodeficiency virus immunogenic peptide

Sequence 4716 BP; 1041 A; 1684 C; 1463 G; 528 T; 0 U; 0 Other;

Query Match 99.9%; Score 1477.4; DB 8; Length 4716;



Best Local Similarity 99.9%; Pred. No. 4,3e-185;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGGCGCCCGCGCAGCATCTGCGCGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60
DB 7 ATGGGCGCCCGCGCAGCATCTGCGCGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 66
QY 61 CTGGCGCCCGCGCGCAAGAGTGTCTATGATGAAACAACCTGGTGTGGGCAACCGCGAG 120
DB 67 CTGGCGCCCGCGCGCAAGAGTGTCTATGATGAAACAACCTGGTGTGGGCAACCGCGAG 126
QY 121 CTGGAAGAGTTGCGCCCTGGAACCCCGGCTGTGGAAGACGAGAGGGCTGCAAGCATC 180
DB 127 CTGGAAGAGTTGCGCCCTGGAACCCCGGCTGTGGAAGACGAGAGGGCTGCAAGCATC 186
QY 181 ATCCGCGAGCTGCAACCCCGCTTGCAAGCCGCGAGGAGGCTGAAGAGCTTTCAAC 240
DB 187 ATCCGCGAGCTGCAACCCCGCTTGCAAGCCGCGAGGAGGCTGAAGAGCTTTCAAC 246
QY 241 ACCGTCGCAACCTGTATCTGCGTGCAGAGAGATGAGGTCCGCGACCAAGAGAGCC 300
DB 247 ACCGTCGCAACCTGTATCTGCGTGCAGAGAGATGAGGTCCGCGACCAAGAGAGCC 306
QY 301 CTGGAAGAGTTGCGCGCGCGCAAGAGTGTGGAAGAGTGTGAGAGAGAGAGAGAG 360
DB 307 CTGGAAGAGTTGCGCGCGCGCAAGAGTGTGGAAGAGTGTGAGAGAGAGAGAGAG 366
QY 361 GCGGCAAGAGGCGAAGGTGAGCCGAACCTACCCCATCTGTGCAAGACTTGCAAGGCG 420
DB 367 GCGGCAAGAGGCGAAGGTGAGCCGAACCTACCCCATCTGTGCAAGACTTGCAAGGCG 426
QY 421 GTGCAACAGAGCCATCAGCCCGCGCACCTCTGGAAGCTGTGGAAGTGTGGAAGAG 480
DB 427 GTGCAACAGAGCCATCAGCCCGCGCACCTCTGGAAGCTGTGGAAGTGTGGAAGAG 486
QY 481 GCCTTCAGCCCGCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCGCACCC 540
DB 487 GCCTTCAGCCCGCGAGGTGATCCCATGTTCAACCGCCCTGAGCGAGGCGCGCACCC 546
QY 541 GACCTGAACAGATGTTGAACACCGTGGGCGGCACAGGCGCGCATGCAAGATGCTGA 600
DB 547 GACCTGAACAGATGTTGAACACCGTGGGCGGCACAGGCGCGCATGCAAGATGCTGA 606
QY 601 GACACCATCAAGAGAGGCGCGCGAGTGGAGCCGCTGCAACCCCGTGCACCGCGGCG 660
DB 607 GACACCATCAAGAGAGGCGCGCGAGTGGAGCCGCTGCAACCCCGTGCACCGCGGCG 666
QY 661 ATGCGCCCGCGCGAGATGCGAGAGCCCGCGCGAGGAGCATGCGCGCGCACACAG 720
DB 667 ATGCGCCCGCGCGAGATGCGAGAGCCCGCGCGAGGAGCATGCGCGCGCACACAG 726
QY 721 CTGGAAGAGATGCGCTGTGATGACCAAGACCCCGCATCCCGTGGGCGCATCTAC 780
DB 727 CTGGAAGAGATGCGCTGTGATGACCAAGACCCCGCATCCCGTGGGCGCATCTAC 786
QY 781 AAGCGGTGATCATCTGGGCGCTGAAACAAGATGTGCGGATGTACAGCCCGTGA 840
DB 787 AAGCGGTGATCATCTGGGCGCTGAAACAAGATGTGCGGATGTACAGCCCGTGA 846
QY 841 CTGGAATCAAGAGAGGCGCGCAAGAGCCCTTCCGCGCATAGTGAACCGCTTCAAG 900
DB 847 CTGGAATCAAGAGAGGCGCGCAAGAGCCCTTCCGCGCATAGTGAACCGCTTCAAG 906
QY 901 ACCCTGCGCGCGAGAGAGCAACCAAGAGGTGAAGACTGATGACCGACACCTGCTG 960
DB 907 ACCCTGCGCGCGAGAGAGCAACCAAGAGGTGAAGACTGATGACCGACACCTGCTG 966
QY 961 GTGCAAGAGCGCAACCCCGCATGCAAGACCATCTGCGGCTTCCGCGCGCGCAC 1020
DB 967 GTGCAAGAGCGCAACCCCGCATGCAAGACCATCTGCGGCTTCCGCGCGCGCAC 1026
QY 1021 CTGGAAGAGATGATGACCGCTGCGAGGCGGTGGGCGGCCCGCACAGGCGCGCTG 1080
```

```
DB 1027 CTGGAAGAGATGATGACCGCTGCGAGGCGGTGGGCGGCCCGCACCAAGCGCGCGT 1086
QY 1081 CTGGCCGAGGCGCATGAGCCAGGCGCAACACCAAGCTGATGATGCAAGAGCAACTTCAG 1140
DB 1087 CTGGCCGAGGCGCATGAGCCAGGCGCAACACCAAGCTGATGATGCAAGAGCAACTTCAG 1146
QY 1141 GGGCGCGCGCGCATGCTCAAGTGTCTCAACTGCGGCAAGAGGCGCATGCGCGCGCAAC 1200
DB 1147 GGGCGCGCGCGCATGCTCAAGTGTCTCAACTGCGGCAAGAGGCGCATGCGCGCGCAAC 1206
QY 1201 TGCCTGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGTGCGGCAAGAGGCGCATGATGA 1260
DB 1207 TGCCTGCGCGCGCGCGCAAGAGGCGCTGTGAAGTGTGCGGCAAGAGGCGCATGATGA 1266
QY 1261 GACTGCAACGAGCGCGCAAGGCGCAACTTCTGTGGCAAGATTTGGCCCAAGCGCAAGGCGCG 1320
DB 1267 GACTGCAACGAGCGCGCAAGGCGCAACTTCTGTGGCAAGATTTGGCCCAAGCGCGCG 1326
QY 1321 CCGGCAACTTCTGTGAGAGCGCGCGCGAGCCCAACCGCGCGCGCGAGAGCTTCCGC 1380
DB 1327 CCGGCAACTTCTGTGAGAGCGCGCGCGAGCCCAACCGCGCGCGCGAGAGCTTCCGC 1386
QY 1381 TTGGAAGAGACCAACCCCGCGCGCAAGAGCAAGAGAGCAAGAGAGAGAGAGAGAG 1440
DB 1387 TTGGAAGAGACCAACCCCGCGCGCAAGAGCAAGAGAGCAAGAGAGAGAGAGAGAG 1446
QY 1441 CTGAAGAGCTGTGTGGCAAGCAACCCCTGTAGCCAGTAA 1479
DB 1447 CTGAAGAGCTGTGTGGCAAGCAACCCCTGTAGCCAGTAA 1485

RESULT 12
ADCl3238
ID ADCl3238 standard; DNA; 4716 BP.
XX
AC ADCl3238;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagProteinRevTatRevNef_C SEQ ID NO 17.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
XX
PR 31-AUG-2001; 2001US-0316860P.
XX
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megele J, Barnett SM, Lian Y, Engelbrecht S, Van Renaburg EJ,
XX
DR WPI; 2003-221593/21.
XX
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
XX Prot, or Rev polypeptide, useful for immunisation, or generating
XX packaging cell lines.
XX
XX Disclosure; Fig 14; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
```

expression cassette can be used to treat HIV type C by gene therapy or used in the development of a vaccine. The gene delivery vector is administered intramuscularly, intradermally, intranasally, subcutaneously, intradermally, transdermally, intravenously, intrarectally, orally or intravenously. The expression cassette is useful for immunisation, generating packaging cell lines and producing HIV polyepitopes. This polynucleotide sequence represents the DNA of an HIV Type C related sequence of the invention.

Sequence 4716 BP; 1041 A; 1684 C; 1463 G; 528 T; 0 U; 0 Other;

Query Match 99.9%; Score 1477.4; DB 10; Length 4716;  
Best Local Similarity 99.9%; Pred. No. 4.3e-185;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGGCGCCGCGCGAGCATCTCTGCGGCGGCGAAGCTGAGCGCTGGAGGCGCATCCGC 60
DB 7 ATGGGCGCCGCGCGAGCATCTCTGCGGCGGCGAAGCTGAGCGCTGGAGGCGCATCCGC 66
QY 61 CTGGCGCCCGCGCGGCGAAGAGTCTACATGATGAGACACTGTGTGGGCGAGCGCGAG 120
DB 67 CTGGCGCCCGCGGCGAAGAGTCTACATGATGAGACACTGTGTGGGCGAGCGCGAG 126
QY 121 CTGGAAGATTGCGCCCTGAAACCCCGGCTGTGAGAGACGAGGAGGCTGCAAGAGATC 180
DB 127 CTGGAAGATTGCGCCCTGAAACCCCGGCTGTGAGAGACGAGGAGGCTGCAAGAGATC 186
QY 181 ATCCGCGAGCTGACCCCGGCTGTGAGAGACGAGGAGGAGCTGAGAGCTGTTCAC 240
DB 187 ATCCGCGAGCTGACCCCGGCTGTGAGAGACGAGGAGGAGCTGAGAGCTGTTCAC 246
QY 241 ACCGCGCGAGCTGACCTGACCTGAGAGAGATGAGAGTCCGAGACACCAAGAGGCG 300
DB 247 ACCGCGCGAGCTGACCTGACCTGAGAGAGATGAGAGTCCGAGACACCAAGAGGCG 306
QY 301 CTGGAAGATTGAGAGAGAGAGAGAGAGAGAGAGAGAGATCAGAGAGGCGAGGCG 360
DB 307 CTGGAAGATTGAGAGAGAGAGAGAGAGAGAGAGAGAGATCAGAGAGGCGAGGCG 366
QY 361 GCGGCAAGAGGCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 367 GCGGCAAGAGGCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
QY 421 GTGCAACAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 427 GTGCAACAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486
QY 481 GCGTTCAAGCCCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 487 GCGTTCAAGCCCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546
QY 541 GACCTGAACAGATGTTGAACAGCGTGGGCGGCGACAGGCGCGCATGAGTGTGAAG 600
DB 547 GACCTGAACAGATGTTGAACAGCGTGGGCGGCGACAGGCGCGCATGAGTGTGAAG 606
QY 601 GACACCATCAACAGAGAGAGGCGCGAGTGGAGAGCGCGTGCACCCGTGAGAGCGGCGCC 660
DB 607 GACACCATCAACAGAGAGAGGCGCGAGTGGAGAGCGCGTGCACCCGTGAGAGCGGCGCC 666
QY 661 ATGCGCCCGGCGAGATGCGAGAGGCGCGGCGAGAGAGATGCGCGGAGAGAGAGAG 720
DB 667 ATGCGCCCGGCGAGATGCGAGAGGCGCGGCGAGAGAGATGCGCGGAGAGAGAGAG 726
QY 721 CTGCAAGAGAGAGATGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 727 CTGCAAGAGAGAGATGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
QY 781 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGAGATGACGCGCGTGAGCATC 840
DB 787 AAGCGGTGATCATCTGGGCGCTGAACAAGATCGTGGAGATGACGCGCGTGAGCATC 846
QY 841 CTGACATCAAGAGAGAGGCGCAAGAGAGCGCTTCCGAGCTACGTGAGCGGCTTTCAAG 900
```

```
DB 847 CTGACATCAAGAGAGAGGCGCCCAAGAGAGCGCTTCCGAGCTACGTGAGAGCGCTTTCAAG 906
QY 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 907 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966
QY 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 967 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
QY 1021 CTGGAAGAGATGATGACCGGCTGCGAGAGGCGGTGGGCGGCGCCAGAGAGAGAGAGAG 1080
DB 1027 CTGGAAGAGATGATGACCGGCTGCGAGAGGCGGTGGGCGGCGCCAGAGAGAGAGAGAG 1086
QY 1081 CTGGCGAGGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1087 CTGGCGAGGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146
QY 1141 GCGCCCGCGCGAGTGTGATGCTTCACTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1147 GCGCCCGCGCGAGTGTGATGCTTCACTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
QY 1201 TGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1207 TGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
QY 1261 GACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1267 GACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
QY 1321 CCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1327 CCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1386
QY 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1387 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
QY 1441 CTGAAGAGCGTGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479
DB 1447 CTGAAGAGCGTGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1485

RESULT 13
ACA03524
ID ACA03524 standard; DNA; 2742 BP.
XX
AC ACA03524;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA encoding immunogenic HIV peptide #7.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
MO WO2003004657-A1.
XX
PN 16-JAN-2003.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
XX
PR 31-AUG-2001; 2001US-0316860P.
XX
PR 16-JAN-2002; 2002US-0349728P.
XX
PR 16-JAN-2002; 2002US-0349739P.
XX
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR ) CHIRON CORP.
XX
```



PN WO2003004620-A2.  
XX  
PD 16-JAN-2003.  
XX  
XX 05-JUL-2002; 2002WO-US021420.  
XX  
PF 05-JUL-2001; 2001US-0303192P.  
XX PR 31-AUG-2001; 2001US-0316860P.  
XX PR 16-JAN-2002; 2002US-0349871P.  
XX  
XX (CHIR) CHIRON CORP.  
XX (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megele J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg BJ;  
XX MPI; 2003-221593/21.  
XX  
XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15NaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
XX packaging cell lines.  
XX  
PS Disclosure; Fig 17; 301pp; English.  
XX  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15NaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC administered intramuscularly, intravenously, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunization, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
XX Type C related sequence of the invention.  
XX  
SQ Sequence 2742 BP; 578 A; 1020 C; 859 G; 285 T; 0 U; 0 Other;

Query Match 99.8%; Score 1475.8; DB 10; Length 2742;  
Best Local Similarity 99.9%; Pred No. 7.5e-185;  
Matches 1477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGCGCCCGGCGAGCATCTGCGCGCGCGGCAAGTGAAGCCTTGGAGGCGATCCGC 60  
DB 7 ATGGGCGCCCGGCGAGCATCTGCGCGCGCGGCAAGTGAAGCCTTGGAGGCGATCCGC 66  
QY 61 CTGGCGCCCGGCGGGAAGAAGTGTACATGATGAAGCACTGTGTGTGGCCAGCCGCGAG 120  
DB 67 CTGGCGCCCGGCGGGAAGAAGTGTACATGATGAAGCACTGTGTGTGGCCAGCCGCGAG 126  
QY 121 CTGGAGAGTTCGCTCTGAACCCCGGCTGTGAGACGAGCGAGGCTGCAAGCAGATC 180  
DB 127 CTGGAGAGTTCGCTCTGAACCCCGGCTGTGAGACGAGCGAGGCTGCAAGCAGATC 186  
QY 181 ATTCGCGAGCTGCAACCCCGCTTGCAGACCGGCAAGGAGCTGAAGAGCCTTGTCAAC 240  
DB 187 ATTCGCGAGCTGCAACCCCGCTTGCAGACCGGCAAGGAGCTGAAGAGCCTTGTCAAC 246  
QY 241 ACCGTGGCCACCTCTACTATGCTGTGACGAGAAAGATTCGCGGACACCAAGAGGCC 300  
DB 247 ACCGTGGCCACCTCTACTATGCTGTGACGAGAAAGATTCGCGGACACCAAGAGGCC 306  
QY 301 CTGGCAAGATCGAGAGGAGGAGCAAGAACTAGTCCAGAGAGATCAGCAGGCGAGGCC 360  
DB 307 CTGGCAAGATCGAGAGGAGGAGCAAGAACTAGTCCAGAGAGATCAGCAGGCGAGGCC 366  
QY 361 GCCGCAAGGCGCAAGGTGAGCAAGAACTAACCTCTGTGCAAGAACTTGCAGGCGCAGATG 420  
DB 367 GCCGCAAGGCGCAAGGTGAGCAAGAACTAACCTCTGTGCAAGAACTTGCAGGCGCAGATG 426  
QY 421 GTGACCAAGGCGCATCAGGCGCCCGGCACTTGAAGCCTTGGGTGAAGGTATGAGAGAGAG 480  
DB 427 GTGACCAAGGCGCATCAGGCGCCCGGCACTTGAAGCCTTGGGTGAAGGTATGAGAGAGAG 486

QY 481 GCCTTCAAGCCCGGAGGTATCCCATGTTCAACCGGCTTGAGCGAGGCGCCACCCCGCAG 540  
DB 487 GCCTTCAAGCCCGGAGGTATCCCATGTTCAACCGGCTTGAGCGAGGCGCCACCCCGCAG 546  
QY 541 GACCTGAACAGATGTTGAACACCGTGGGCGGCGCACCGGCGGCGCATGACAGTGTGAAG 600  
DB 547 GACCTGAACAGATGTTGAACACCGTGGGCGGCGCACCGGCGGCGCATGACAGTGTGAAG 606  
QY 601 GACACATCAACGAGAGAGCGCGGAGTGGGACCGCGGTGACACCCCGTGCACGCGCGGCC 660  
DB 607 GACACATCAACGAGAGAGCGCGGAGTGGGACCGCGGTGACACCCCGTGCACGCGCGGCC 666  
QY 661 ATTCGCCCCCGGCGAGTATGCGGAGACCCCGGCGGCGGACATCGCGGCGCACCAAGCAC 720  
DB 667 ATTCGCCCCCGGCGAGTATGCGGAGACCCCGGCGGCGGACATCGCGGCGCACCAAGCAC 726  
QY 721 CTGCAAGAGCAGATCGCTGATGACCAACACCCCGCATCCCGTGGGCGCATCTAC 780  
DB 727 CTGCAAGAGCAGATCGCTGATGACCAACACCCCGCATCCCGTGGGCGCATCTAC 786  
QY 781 AAGCGTGTGATCATCTTGGGCTTGAACAGATCGTGGGATGTACAGCCCGTGAGCATC 840  
DB 787 AAGCGTGTGATCATCTTGGGCTTGAACAGATCGTGGGATGTACAGCCCGTGAGCATC 846  
QY 841 CTGGAACATCAAGAGGCGCCCGAAGAGCCCTTCCGCGCATACGTGACCGCTTTTCAAG 900  
DB 847 CTGGAACATCAAGAGGCGCCCGAAGAGCCCTTCCGCGCATACGTGACCGCTTTTCAAG 906  
QY 901 ACCCTGGCGCGCGAGAGAGCACCAAGAGGTGAAGAATGATGACCGACACCCCTGCTG 960  
DB 907 ACCCTGGCGCGCGAGAGAGCACCAAGAGGTGAAGAATGATGACCGACACCCCTGCTG 966  
QY 961 GTGCAAGACGCAACCCCGATGCAAGACCAATCTGCGCGCTTGGCGCCCGGCGCAGC 1020  
DB 967 GTGCAAGACGCAACCCCGATGCAAGACCAATCTGCGCGCTTGGCGCCCGGCGCAGC 1026  
QY 1021 CTGGAAGATGATGAACCGCTGCGAGGCGGTGGCGCGCCCGACCAAGGCGCGCGTG 1080  
DB 1027 CTGGAAGATGATGAACCGCTGCGAGGCGGTGGCGCGCCCGACCAAGGCGCGCGTG 1086  
QY 1081 CTGCGCAGAGCGATGAGCCAGGCGCAACACAGCGTGTATGTGAGAGAGCAACTTCAAG 1140  
DB 1087 CTGCGCAGAGCGATGAGCCAGGCGCAACACAGCGTGTATGTGAGAGAGCAACTTCAAG 1146  
QY 1141 GGCCCCCGGCGCATCTGTCAGATGCTTCAATCGGCGCAAGAGGCGCCATCGCCGCAAC 1200  
DB 1147 GGCCCCCGGCGCATCTGTCAGATGCTTCAATCGGCGCAAGAGGCGCCATCGCCGCAAC 1206  
QY 1201 TGCGGCGCCCCCGCAAGAGGCGCTGTGAAGTGGCGCAAGAGGCGCCACGATGAAG 1260  
DB 1207 TGCGGCGCCCCCGCAAGAGGCGCTGTGAAGTGGCGCAAGAGGCGCCACGATGAAG 1266  
QY 1261 GACTGCAACGAGCGCCAGGCGCAACTTCTTGGGCAAGATCTTGGCCAGCCCAAGAGGCGCG 1320  
DB 1267 GACTGCAACGAGCGCCAGGCGCAACTTCTTGGGCAAGATCTTGGCCAGCCCAAGAGGCGCG 1326  
QY 1321 CCGGCAACTTCTTGTGAGAGCGCGCGCGGAGCCCAAGGCGCCCGCGGAGGCTTCGCG 1380  
DB 1327 CCGGCAACTTCTTGTGAGAGCGCGCGCGGAGCCCAAGGCGCCCGCGGAGGCTTCGCG 1386  
QY 1381 TTGAGAGAGCAACCCCGGCGCAAGAGGAGAGAGCAAGGACCGCGAGACCTTGAACAGC 1440  
DB 1387 TTGAGAGAGCAACCCCGGCGCAAGAGGAGAGAGCAAGGACCGCGAGACCTTGAACAGC 1446  
QY 1441 CTGAAGAGCTGTTCGCAACGACCCCTTGAAGCAGTAA 1479  
DB 1447 CTGAAGAGCTGTTCGCAACGACCCCTTGAAGCAGTAA 1485

RESULT 15  
ADCI3230  
ID ADCI3230 standard; DNA; 3930 BP.

XX AC ADCl3230;  
XX 18-DEC-2003 (first entry)  
DE DNA of HIV construct GagComp1Polmut\_C SEQ ID NO 9.  
XX expression cassette; HIV Gag; Env; Int; Nef; p15RaseH; Pol; Tat; Prox;  
KM Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX Human immunodeficiency virus.  
OS MO2003004620-A2.  
XX PN 16-JAN-2003.  
XX PF 05-JUL-2002; 2002MO-US021420.  
XX PR 05-JUL-2001; 2001US-0303192P.  
XX PR 31-AUG-2001; 2001US-0316660P.  
XX PR 16-JAN-2002; 2002US-0349871P.  
XX PA (CHIR ) CHIRON CORP.  
XX PA (UYST-) UNIV STELLENBOSCH.  
XX PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ,  
XX DR WPI, 2003-221593/21.  
XX PT New expression cassette comprising a polynucleotide sequence encoding a  
XX PT polypeptide including an HIV Gag, Env, Int, Nef, p15RaseH, Pol, Tat,  
XX PT Prox, or Rev polypeptide, useful for immunization, or generating  
XX PT packaging cell lines.  
XX PS Disclosure; Fig 6; 301pp; English.  
XX CC The invention relates to a novel expression cassette comprising a  
XX CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
XX CC Int, Nef, p15RaseH, Pol, Tat, Prox, or Rev polypeptide. The novel  
XX CC expression cassette can be used to treat HIV type C by gene therapy or  
XX CC used in the development of a vaccine. The gene delivery vector is  
XX CC administered intramuscularly, intravenously, intranasally,  
XX CC subcutaneously, intradermally, transdermally, intravaginally,  
XX CC intrarectally, orally or intravenously. The expression cassette is useful  
XX CC for immunisation, generating packaging cell lines and producing HIV  
XX CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
XX CC Type C related sequence of the invention.  
XX SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;  
Query Match 99.8%; Score 1475.8; DB 10; Length 3930;  
Best Local Similarity 99.9%; Pred. No. 7.1e-185;  
Matches 1477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATGGGCGCCGCGCAGCATCTGCGCGCGGCAAGCTGGAACGCTGGAGCGCATCCG 60  
DB 7 ATGGGCGCCGCGCAGCATCTGCGCGCGGCAAGCTGGAACGCTGGAGCGCATCCG 66  
QY 61 CTGCGCGCCGCGCGCAAGAGTCTACATGATGAAGCACTGTGTGTGGCCGCGAG 120  
DB 67 CTGCGCGCCGCGCGCAAGAGTCTACATGATGAAGCACTGTGTGTGGCCGCGAG 126  
QY 121 CTGGAAGATTGGCTTGAACCCCGGCTGTGGAACCAAGCGAGGCTGGCAAGCATC 180  
DB 127 CTGGAAGATTGGCTTGAACCCCGGCTGTGGAACCAAGCGAGGCTGGCAAGCATC 186  
QY 181 ATCCGCGAGCTGACCCCGGCTGTGGAACCGGCGAGGAGCTGAAGAGCTTTCAAC 240  
DB 187 ATCCGCGAGCTGACCCCGGCTGTGGAACCGGCGAGGAGCTTTGAAGCTTTTCAAC 246  
QY 241 ACCGTGGCAACCTGTACTGCGTGAAGAGAGATGAGGTCCGCAACCAAGAGGCC 300  
DB 247 ACCGTGGCAACCTGTACTGCGTGAAGAGAGATGAGGTCCGCAACCAAGAGGCC 306

QY 301 CTGGAAGATGAGGAGGAGCAAGAGTGCAGAGATTCAGAGCGCGAGGCC 360  
DB 307 CTGGAAGATGAGGAGGAGCAAGAGTGCAGAGATTCAGAGCGCGAGGCC 366  
QY 361 GCCGCAAGGAGCAAGTGAAGCCAGACTACCCCATGTGCAGAACCTGCAGGCGCAGATG 420  
DB 367 GCCGCAAGGAGCAAGTGAAGCCAGACTACCCCATGTGCAGAACCTGCAGGCGCAGATG 426  
QY 421 GTGCACCAAGGCAATCAGCCCGGCAACCTGGAACGCTGTGAGTGAAGTATCGAGAGAG 480  
DB 427 GTGCACCAAGGCAATCAGCCCGGCAACCTGGAACGCTGTGAGTGAAGTATCGAGAGAG 486  
QY 481 GCCTTACGCGCGAGGATGATCCCATGTTCAACCGGCTGAGGAGGCGCCACCCCGAG 540  
DB 487 GCCTTACGCGCGAGGATGATCCCATGTTCAACCGGCTGAGGAGGCGCCACCCCGAG 546  
QY 541 GACCTGAACACGATGTTGAACACCGTGGCGGCAACGAGCGGCAATGCAAGTCTGAAG 600  
DB 547 GACCTGAACACGATGTTGAACACCGTGGCGGCAACGAGCGGCAATGCAAGTCTGAAG 606  
QY 601 GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGACACCCGTCGACCGCGGCCCC 660  
DB 607 GACACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGACACCCGTCGACCGCGGCCCC 666  
QY 661 ATGCGCCCGGCGCAAGTGGCGAGCCCGGCGCAAGATGCGCGGCAACCAAGCACC 720  
DB 667 ATGCGCCCGGCGCAAGTGGCGAGCCCGGCGCAAGATGCGCGGCAACCAAGCACC 726  
QY 721 CTGCAAGACAGATCGCTGATGACCAAGCAACCCCGCATCCCTGGGCGCATCTAC 780  
DB 727 CTGCAAGACAGATCGCTGATGACCAAGCAACCCCGCATCCCTGGGCGCATCTAC 786  
QY 781 AAGCGGTGATCATCTGCGCTGAAACAAGATCTGCGAGTGAACGCCCTGTAGCATC 840  
DB 787 AAGCGGTGATCATCTGCGCTGAAACAAGATCTGCGAGTGAACGCCCTGTAGCATC 846  
QY 841 CTGGAATCAACAGGCGCGCAAGAGCCCTTCCGGAATCTAGTGAACCGCTTCTTCAAG 900  
DB 847 CTGGAATCAACAGGCGCGCAAGAGCCCTTCCGGAATCTAGTGAACCGCTTCTTCAAG 906  
QY 901 ACCCTGCGCGCGAGCAAGCAACCAAGAGTGAAGACTGATGACCAACCTGCTG 960  
DB 907 ACCCTGCGCGCGAGCAAGCAACCAAGAGTGAAGACTGATGACCAACCTGCTG 966  
QY 961 GTGCAGAACGCCAACCCCGCATCTGCAAGACCATCTGCGGCTCTGCGCCCGCGCAC 1020  
DB 967 GTGCAGAACGCCAACCCCGCATCTGCAAGACCATCTGCGGCTCTGCGCCCGCGCAC 1026  
QY 1021 CTGGAAGATGATGACCGCTGCGCAAGGCGCTGCGCGCGCCCAAGCAAGGCGCGCTG 1080  
DB 1027 CTGGAAGATGATGACCGCTGCGCAAGGCGCTGCGCGCGCCCAAGCAAGGCGCGCTG 1086  
QY 1081 CTGCGCGAGCGATGACCGAGCAACCAAGCTGATGATGACAGAGAGCAACTTCAAG 1140  
DB 1087 CTGCGCGAGCGATGACCGAGCAACCAAGCTGATGATGACAGAGAGCAACTTCAAG 1146  
QY 1141 GCGCCCGGCGCATCTGCAAGTCTTCAACTGCGGCAAGAGGCGCCCATCTGCGCAAC 1200  
DB 1147 GCGCCCGGCGCATCTGCAAGTCTTCAACTGCGGCAAGAGGCGCCCATCTGCGCAAC 1206  
QY 1201 TGCAGCGCGCGCGCAAGAGGCTGTGGAAGTGTGCGGCAAGAGGCGCCCAAGATGAAG 1260  
DB 1207 TGCAGCGCGCGCGCAAGAGGCTGTGGAAGTGTGCGGCAAGAGGCGCCCAAGATGAAG 1266  
QY 1261 GACTGACAGAGCGCAAGGCGCAACTTCTGTGCAAGATCTGCGGCAAGAGGCGCGC 1320  
DB 1267 GACTGACAGAGCGCGCAAGGCGCAACTTCTGTGCAAGATCTGCGGCAAGAGGCGCGC 1326  
QY 1321 CCGGCAACTTCTGTGCAAGGCGCGCGCAAGGCGCGCGCGCGCGGAGGCTTCCGC 1380  
DB 1327 CCGGCAACTTCTGTGCAAGGCGCGCGCAAGGCGCGCGCGCGCGGAGGCTTCCGC 1386

Qy	1381	TTGAGAGACCA	CCCCCGGCCA	GAGAGAGAGCA	CGGAGACCTGA	CCAGC	1440
Db	1387	TTGAGAGACCA	CCCCCGGCCA	GAGAGAGAGCA	CGGAGACCTGA	CCAGC	1446
Qy	1441	CTGAGAGCCTG	TTGGCAACGA	CCCCCTGAGC	CAATA		1479
Db	1447	CTGAGAGCCTG	TTGGCAACGA	CCCCCTGAGC	CAATA		1485

Search completed: March 12, 2005, 17:35:52  
Job time : 843.981 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2005, 16:34:11 ; Search time 5208.18 Seconds  
(without alignments)  
10809.355 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 1479

Sequence: 1 atggggcggccgcggccagcatc.....acgacccctgagccagtaaa 1479

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:

1: gb\_ests1: \*  
2: gb\_ests2: \*  
3: gb\_hnc: \*  
4: gb\_ests3: \*  
5: gb\_ests4: \*  
6: gb\_ests5: \*  
7: gb\_ests6: \*  
8: gb\_g881: \*  
9: gb\_g882: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	82.6	5.6	746	7	CK215205 FGAS02715
2	81.2	5.5	749	5	BQ744212 WHE4113_A
3	80	5.4	1064	7	CK163121 FGAS01573
4	80	5.4	2277	9	CL968487 OsIFCC040
5	79.8	5.4	1126	7	CK211108 FGAS02294
6	79.4	5.4	826	7	CN128942 RH0H1_32
7	78	5.3	807	6	CB673622 OSUNE06D
8	78	5.3	821	6	CB662794 OSUNE07C
9	78	5.3	824	6	CB666192 OSUNE07H
10	78	5.3	851	6	CB645622 OSUNE07H
11	76.6	5.2	807	6	CB673355 OSUNE07N
12	76.6	5.2	1138	7	CK163513 FGAS01614
13	76.4	5.2	812	6	CB681973 OSUNE108M
14	76.4	5.2	832	6	CB684834 OSUNE114L
15	76.2	5.2	759	2	BF259495 HVSMF001
16	76.2	5.2	1173	9	CL979928 OsIFCC045
17	75.6	5.1	562	2	BF484304 WHE2321_F
18	75.6	5.1	649	6	CB870888 HCL5T04W
19	75.4	5.1	834	6	CB878700 OSUNE07C
20	75.4	5.1	840	6	CB873814 OSUNE08I
21	74.8	5.1	766	6	CB629976 OSIEB06E
22	74.8	5.1	797	6	CB658304 OSIEB14E
23	74.8	5.1	810	6	CB618374 OSIEB02D
24	74.8	5.1	823	6	CB627740 OSIEB02C

25	74.8	5.1	826	6	CB641789 OSUNE01J
26	74.8	5.1	838	6	CB659169 OSUNE15O
27	74.8	5.1	865	6	CB628660 OSIEB04P
28	74.6	5.0	708	6	CB878344 AZ04_1021
29	74.6	5.0	818	6	CB652454 OSUNE02G
30	74.4	5.0	1962	9	CL961326 OsIFCC005
31	74.2	5.0	785	4	BG369139 HVSMF1002
32	74.2	5.0	933	8	CC391778 PHU1370TB
33	73.6	5.0	791	6	CB650382 OSUNE14M
34	73.4	5.0	731	5	BQ752847 WHE4119_G
35	73	4.9	2629	9	CL955065 OsJRU000
36	72.6	4.9	2598	3	AY103647 Zea mays
37	72.4	4.9	648	2	BE517305 WHE0616_A
38	72.2	4.9	803	6	CB633391 OSIEB12D
39	72.2	4.9	1308	3	AY104577 Zea mays
40	72	4.9	1728	9	CL973038 OsIFCC042
41	71.8	4.9	764	5	BQ804819 WHE3559_C
42	71.4	4.8	718	6	CD938268 OV_109123
43	71.4	4.8	718	6	CD938289 OV_109K04
44	71.2	4.8	782	6	CB661708 OSUNE04P
45	71.2	4.8	2517	9	CL972413 OsIFCC022

## ALIGNMENTS

RESULT 1  
LOCUS CK215205 746 bp mRNA linear EST 09-DEC-2003  
DEFINITION FGAS027158 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum  
ACCESSION CK215205  
VERSION CK215205.1 GI:39621309  
KEYWORDS  
SOURCE  
ORGANISM  
Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticaceae; Triticum.  
1 (bases 1 to 746)

REFERENCE  
AUTHORS Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genewein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilson, D.,  
Penniket, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas.ests@usask.ca

FEATURES  
source  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [13,662].  
Plate: L6B016, row: K, column: 15.  
Location/Qualifiers  
1..746  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
/note="Organ: Crown and leaf; Vector: pCMV\_SPORT6; Crown  
(50%) and leaf (50%) tissues from wheat cultivar Norstar  
after short exposure times to low temperature in the light  
and in the dark. 12 mRNA populations were combined before  
constructing the library. The first 6 populations: After 7  
days of growth at 20°C from wheat cultivar Norstar after  
short exposure times to low temperature in the light and

in the dark. 12 mRNA populations were combined before constructing the library. The first 6 populations: After 7 days of growth at 20, wheat plants were transferred to 4C in the light. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. The last 6 populations: After 7 days of growth at 20C, wheat plants were transferred to 4C in the dark. 1cm crown sections and green leaf tissue were separately harvested after 1, 3, and 6 hours of low temperature exposure. First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI. In addition, this library used a primer for second strand synthesis that annealed to an artificial sequence (RNA oligo) added before first strand synthesis. Therefore when sequences from EST generated from this library will be masked for vector and adaptor sequences, an additional masking step will have to be included to mask this RNA oligo that is common to all clones (sequence CGACTGAGACGAGGACCTGACATGACTGAGAGTAGAA). "

## ORIGIN

Query Match 5.6%; Score 82.6; DB 7; Length 746;  
Best Local Similarity 48.4%; Pred. No. 4.6e-05;  
Matches 288; Conservative 0; Mismatches 301; Indels 6; Gaps 2;

```

QY 124 GAGAGTTCGCGCTGAAACCCCGCTGCTGAGACCGAGAGGCTGCAAGATCATC 183
DB 737 GAGATNNCGAAAAACCCCATGAAACCTGATGACGACTCGCGAGCCAGGCGCATTC 678
QY 184 CGCAGCTGCAACCCCGCTGAGACCGGACGAGAGCTGAGAGCTGTTCAACACC 243
DB 677 CTCACATGCTGCTCAAGCTCATCGGCGCCAGAGACCATGAGATGCGCTCTACACC 618
QY 244 GTGGCCACCTCTGATCTGCTGCAAGAAAGATGAGTCCGCGACACCAAGAGCCCTG 303
DB 617 GGGTACTCCCTCTCGCCACCGCGCTCGCCATCCCGGACGAGGACCATCTTGCCATG 558
QY 304 GACAGATGAGAGGAGAGGAAACAAAGTCCGAGGAGAGATCCAGACCGGAGGCGCC 363
DB 557 GACATCAACCGGAGAACTACAGCTGGGGGCTGCGTGATGAGAAAGCGCGCTGCG 498
QY 364 GACAAAGGAGAGTGAAGCCAGAACTACCTGTCAGAACTGCAAGGCGCAGATGCTG 423
DB 497 CACAAAGATGCACTTCGCGAGGG---CCGCGGCTGCGGCTGCGAGCGCTGCTGAG 441
QY 424 CACCGAGCCATAGCCCGGACACCTGAAAGCTGGGTGAAGTGAAGTGAAGAGAGCC 483
DB 440 GACGAGGCGCAACGAGGACCTTTCGACTTCCTTGTGACGCGCAGACAACTAC 381
QY 484 TTCAGCCCCGAGGTATCCCATGTTCAACCGCCCTGAGAGGAGGCGCAACCCCGAGAG 543
DB 380 CTCACATCAACGAGAGCGCTCATGAAAGCTGTCAGAGCTGCGGCGCTCTCGCTACAG 321
QY 544 CTGAACACAGATGTTGAACACCGTGGGCGGCGCACGAGCGCCATGACAGATGCTGAG 603
DB 320 ---AACGAGCTTGAAGAGGCTCGGTGAGTCCCGCGCGCGCCCATGAGGAGAGTAC 264
QY 604 ACCATCAACGAGAGAGCGCGGAGTGGAGACCGCGTGCACCCCGTGCAGCGCGGCCATC 663
DB 263 ATCCGCTACTACCGGAGCTTGTCTTCGACTTCACCAAGAGCGCTTCGCGCGAGCCAG 204
QY 664 GCCCGGCGCAGATGCGGAGCGCGCGGAGCGGACATGCGCGGACCAACAGCA 718
DB 203 GTCCAGATCTGCGAGCTCCCGTGGCGAGCGGATCACTCTTGTGCGCGGCGCA 149

```

RESULT 2  
LOCUS B0744212  
DEFINITION WHE4113\_A01 A01ZS wheat salt-stressed root cDNA library Triticum aestivum cDNA clone WHE4113\_A01, mRNA sequence.  
ACCESSION B0744212

VERSION B0744212.1 GI:21890999  
KEYWORDS EST  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.  
1 (bases 1 to 749)  
AUTHORS Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K., Dvorak,J., Izzo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.  
TITLE The structure and function of the expressed portion of the wheat genomes - Salt-stressed root cDNA library  
JOURNAL Unpublished (2002)  
COMMENT Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@wr.usda.gov  
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
Seq primer: SK primer.  
FEATURES  
source  
1..749  
location/Qualifiers  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE4113\_A01 A01"  
/tissue\_type="Roots"  
/dev\_stage="Full tillering"  
/lab\_host="E. coli SOLR"  
/clone\_id="Wheat salt-stressed root cDNA library"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid plasmid SK(-); Site 1: EcoRI; Site 2: XhoI; Hydropic plants grown to full tillering stage were treated with 150 mM NaCl for either 12 hours or 7 days. Root tissues of the plants subjected to both types of treatment were collected separately at University of California, Davis (E. Akhunov and K. Deal in J. Dvorak's lab). Total RNA was prepared separately from the two samples (12h and 7day treatments), and equal amount of RNA was then pooled. PolyA RNA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give Bluescript SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak) at the University of California, Davis. Colony plating, plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

## ORIGIN

Query Match 5.5%; Score 81.2; DB 5; Length 749;  
Best Local Similarity 47.5%; Pred. No. 8.4e-05;  
Matches 282; Conservative 0; Mismatches 303; Indels 9; Gaps 1;

```

QY 280 GTCCGAGACCAAGAGGCGCTGGAACAATGAGAGGAGAGCAAGATGCGACAG 339
DB 19 GTGACGCCAAGCGCGCGCCAGACACCGCGCTTGGGCGCAAGTGGCGCTTCAACAG 78
QY 340 AAGATCCAGAGGCGCGAGCGCGCGACAGAGGAGGAGTGAAGCACTACCCCATGTG 399
DB 79 CGGCGCCAGCTCCCGCCAGCGCGCGGAGAGAGTGAAGGCGCAAGCGCGAGGCGGAG 138
QY 400 CAGAACTGCAAGGCGCAGATGTGCACAGGCGCATAGCCCCCGACCTGAAAGCTTG 459
DB 139 AAGGAGCGCGCGCGCGCGAGCGCGCTACAGAAACCAACCGCTTCCACACACAGGGC 198
QY 460 GTGAAGGTGATGAGAGAGAGGCGCTTCAAGCGCGAGGTGATCCCATGTTTCAACCGCCTG 519
DB 199 TTCCGCGGCAACAGCAAGAAACACCGCAGAGCTTCCGTGCACATTCACGCGGACTGAGC 258
QY 520 AGCGAGGCGCGACCCCGCAGAGCTGAAACAGATGTTGAACAACGTGGGCGGCGACAG 579

```



Db 259 ATCTGTGAGAGATGATCCCTTTCGCAATTGACCAAGCTCTTTCGCGCGTCCGACGAC 318  
Qy 580 GCCGCATGATGATGCTGAAAGACACCATATCAAGAGAGGCGCCGATGAGAGCCGCG 639  
Db 319 CCCGAGACCTGCTCTGCGCGCGCTGACCTGACGACCGCCCTACGACCCGCTC 378  
Qy 640 CACCCGTGACGCGCGCCCATGCGCCCGGCGAGATGCGAGGCGCGCGACGAC 699  
Db 379 AACCCCAAGAGCGCGCGCGCTGAGCGCTTCAAGAACCGCAGTTCTTCAAGATACC 438  
Qy 700 ATGCGCGGACGACGACGACCTGTC-----AGAGAGATGCGCTGATGACGAC 750  
Db 439 ACCACGAGAGAGCCCATATGCGCGCGCTGCGCGAGAGAGAGAGAGAGAGAGAGAG 498  
Qy 751 AACCCCGCATGCGCGCGCGGACATCTACAGCGGTGATCATCTGCGCGCTGACGAC 810  
Db 499 ACCGAGCGCATCTGCGCGCGCGCTGATGACGCGCGCGAGATCTCTCTGCGGACAT 558  
Qy 811 ATGCGCGGATGACGACGCGCGGATGCTGAGATCATCAAGAGAGAGAGAGAGAGAG 864  
Db 559 GTGTGAGAGCGCTGCGCAAGCTCTTCTTTCGACAGCGCGAGGCTCCAG 612

RESULT 3  
CK163121 1064 bp mRNA linear EST 05-DEC-2003  
LOCUS CK163121 Triticum aestivum FGAS: Library 4 Gate 8 Triticum  
DEFINITION FGAS015739 Triticum aestivum FGAS: Library 4 Gate 8 Triticum  
ACCESSION CK163121  
VERSION CK163121.1 GI:38993028  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 1064)  
Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Gemein, B., Grail, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Links, M.G., McCarthy, E.L., Montoy, A., Muzak, I., Nilson, D.,  
Penniket, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgaes@usask.ca

This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [1,834].  
Place: L4B07 row: C column: 11.  
Location/Qualifiers

FEATURES  
source 1..1064  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 4 Gate 8"  
/note="Organ: Crown and leaf; Vector: PCW,SPOR6;  
Conditions for growth: Seeds were germinated in a  
water-saturated mix (50% black earth and 50% ProMix) in a  
growth chamber for 7 days under an irradiance of 200 mmol  
m<sup>-2</sup> sec<sup>-1</sup>. The temperature was maintained at 20 degrees C  
with a 15-hr photoperiod under a relative humidity of 70%.  
After this period watering of plants was stopped. Four  
time points were sampled during a two week period; the  
first after wilting was observed and the last, two weeks  
later, consisted of live crown and leaf tissue (leaf

## ORIGIN

tissue that was yellow was not included in sampled  
material). First strand synthesis in this library was done  
in the presence of methylated dCTP thereby protecting from  
internal cleavage with NotI."

Query Match 5.4%; Score 80; DB 7; Length 1064;

Best Local Similarity 48.4%; Pred. No. 0.00014;  
Matches 283; Conservative 0; Mismatches 296; Indels 6; Gaps 2;

Qy 124 GAGAGTTGCGCCTGTAACCCCGCGCTGAGAGACGAGAGAGGCTGACAGATCATC 183  
Db 282 GAGATCACCGCAACACCATGATGAGAGAGTGGCGAGAGAGGCGCAGTTTC 341  
Qy 184 CGCGAGCTGACCCCGCGCTGACAGACCGGACGAGAGCTGAAGACTTTCACACC 243  
Db 342 CTCACATGCTGCTCAAGTCAATCGCGCGCAAGAGACCATGAGATGCGCTTACAC 401  
Qy 244 GTGGCAGCCCTGTATGCGGTGACAGAGAAAGATCGAGTCCGCGACACCAAGAGGCGCTG 303  
Db 402 GGTACTCTCTCTCTGCGACCGCGCTCGCATCCCGAGACGACCATCTTGGCATG 461  
Qy 304 GACAAATCGAGAGAGAGAGAGAGAGAGAGAGAGATCAGAGATCAGAGCGCGCGC 363  
Db 462 GACATCAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 521  
Qy 364 GACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 423  
Db 522 CACAAAGATGATCTTCCGCGAGAGG---CCGCGCGCTGCGGTCTGACCGCTGCTGAG 578  
Qy 424 CACGAGGCGATGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 483  
Db 579 GAGAGAGCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638  
Qy 484 TTCAGCGCGGAGAGATGATCCCATGTTTCAACCGCGCTGAGAGAGAGAGAGAGAG 543  
Db 639 CTCACATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 698  
Qy 544 CTGAACAGATGTTGAACAGCGTGGCGCGGACAGAGCGCGCATGATGCTGAAGAG 603  
Db 699 ---AACAGCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 755  
Qy 604 ACCATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663  
Db 756 ATCCGCTACTACCGGACTTCTCTCTGACCTTCAAGAGAGAGAGAGAGAGAGAGAG 815  
Qy 664 GCCCGCGCGAGATGCGGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 708  
Db 816 GTGAGATGCTGCGAGCTCCCGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 860

RESULT 4  
CL968487 2277 bp DNA linear GSS 21-SEP-2004  
LOCUS OsrFCC040489 Oryza sativa Expressed Library Oryza sativa (Indica  
DEFINITION cultivar-group) genomic, genomic survey sequence.  
ACCESSION CL968487  
VERSION CL968487.1 GI:52391607  
KEYWORDS GSS.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Euphorbiaceae; Oryzae; Oryza.  
1 (bases 1 to 2277)

REFERENCE Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,  
Wong, G.K.S., Deng, X.W. and Wang, J.  
An analysis of transcriptional regulation of the rice genome and  
its comparison to Arabidopsis  
Unpublished (2004)

TITLE JOURNAL COMMENT  
Department of Bioinformatic

Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481359  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.

# FEATURES

source  
1..2277  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Expressed Library"  
/note="Oryza sativa exon trapped genomic sequences"

## ORIGIN

Query Match 5.4%; Score 80; DB 9; Length 2277;  
Best Local Similarity 42.2%; Pred. No. 0.0015;  
Matches 510; Conservative 0; Mismatches 695; Indels 3; Gaps 1;

5 GGGCCCGCCGACATCTCCGCGGCGCAAGTGGAGCCCTGGAGCGCATCCGCTGC 64  
197 GGACAAACCCAGCGATCTCTTACGCGCGCGCGCGCAACAGGGCGCGCGCGTGA 256  
65 GCGCCGCGCGCAAGAGTCTCATGATGAAGCACTGTGTGGCCAGCGCGAGCTGG 124  
257 TGGCGCGGAGAGCGCTGTCTTCTTACAGACAGCTGTCAAGCGTTTCAACCGCG 316  
125 AGAAGTTGCGCTGAACCCCGCGCTGTGAGAACAGCGAGGCTGCAAGCATATCC 184  
317 TCAGCGCGCGACAGAGCGCGCGCATGGCGAGCTGAGTGTGCGTCAAGCGCTCCG 376  
185 GCCAGCTGACCCCGCGCTGACAGCGCGAGAGAGCTGAAGCTTTCAACCGC 244  
377 ACTCAAGTACACCTTACACACACAGACAGCGCGCGCTGTGCGCATGACACCGC 436  
245 TGGCCACCTGTACTGCTGTCAGAGAGATGAGGTGCGCGACCAAGAGCGCTGG 304  
437 GGAACCGCGCGCTGAGCGCTGTGCGCGAGACATGGCGACCGCTCATCTGCGCG 496  
305 ACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364  
497 ACAAGCGCGCTGACCGCGCGCGCGCGCTGTGCGCGAGAGAGAGAGAGAGAG 556  
365 ACAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424  
557 CCAAGTGGCGCGCGAGTGCAGCTTGTGCGCGCGCGCGCGCAACAAAGCTCATCG 616  
425 ACCAGGCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 484  
617 GCGCGCGCGAGAGCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 676  
485 TCAAGCGCGAGTATCCCATGTTTCAACCGCGCTTGAAGCGAGGCGCGCGCGAG 544  
677 TCGTGAAGGAGCGTCAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 736  
545 ---TGAACAGATGTTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGAG 601  
737 GCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 796  
602 ACACCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661  
797 TCACGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 856  
662 TCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 721  
857 CCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 916  
722 TGCAGAGAGAGATGCGCTGAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 781  
917 TCAATGCGCGCGCTCTTCTTCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 976  
782 AGCGGTGATCATCTTGGGCGCTGAACAAGATGTCGAGATGTAACAGCGCGCGAGCATCC 841

Db 977 TCACCAACTGCGCGCGCTGAGAGCTTACCGTGGCGCGAGACCTGGAGCGCATPA 1036  
Qy 842 TGAACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901  
Db 1037 TCTCCAGAGTCCAGTCCGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1096  
Qy 902 CCTCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 961  
Db 1097 AACCGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1156  
Qy 962 TGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021  
Db 1157 AGCGGTGATCATCTTGGAGCAAGATGCTTGTGATGCGGTGCAAGAGAGAGAG 1216  
Qy 1022 TGAAGAGATGATGACCGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081  
Db 1217 TCGAAGATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1276  
Qy 1082 TGGCGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141  
Db 1277 AGTTCCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1336  
Qy 1142 GCGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1201  
Db 1337 CCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1396  
Qy 1202 GCGCGCGC 1209  
Db 1397 GCTTCGCC 1404

## RESULT 5

### LOCUS

### DEFINITION

### ACCESSION

### VERSION

### KEYWORDS

### ORGANISM

### SOURCE

### REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### COMMENT

CR211108 1126 bp mRNA linear EST 09-DEC-2003  
FGAS022942 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum  
aestivum cDNA, mRNA sequence.  
CR211108  
CR211108.1 GI:39617217  
EST.  
Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.  
1 (bases 1 to 1126)  
Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
Peniket, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas\_est@cs.usask.ca  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [12..750].  
Plate: L6B01 row: A column: 20.

## FEATURES

source  
1..1126  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"  
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown





ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 821)  
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 Large-scale identification of ESTs involved in the interaction  
 between rice and Magnaporthe grisea

TITLE Unpublished (2003)

JOURNAL Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu

COMMENT PCR Primers  
 FORWARD: gta aaa cga cgg cca gtc  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 07 row: C column: 09  
 Seq primer: gta aaa cga cgg cca gtc.  
 Location/Qualifiers  
 1..821  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNBd07C09"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNBd"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
 XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"

ORIGIN

Query Match 5.3%; Score 78; DB 6; Length 821;  
 Best Local Similarity 44.4%; Pred. No. 0.00032;  
 Matches 312; Conservative 0; Mismatches 390; Indels 0; Gaps 0;

204 GGAGACCGGCGAGGAGCTGAAGAGCTGTTCAACACCGTGGCCACCTGATCGCT 263  
 88 GGAGATGAGAAATGGCGCGGAGACGTTCTTCACTCGAGTCCGTGAACGAGGTCA 147  
 264 GCACGAGAAGATCGAGTCCGCGACACCAAGAGGCGCTTGAACAGATCGAGGAGCA 323  
 148 CCGGAGCAAGCTGTGTGACCAAGTGTGGAACGGGTGCTGACGGCTGCTGCCAGGA 207  
 324 GAACAAGTCCAGCAAGATTCAGAGGCCCGGACCGCGACAGGCGCAAGGTGAGCA 383  
 208 CCCCAGCAAGAGTGGCGTGCAGACGTGACCAACCAACATGATGATGTTGTCG 267  
 384 GAACATACCCCATCTGTCAGAACTGCAAGGCCAGATGTGACACAGCCATCAGCCCCG 443  
 268 CGAGATTCACCAACGACCGTGCATCAAGAAATGTCGCGACACCTGCGCGCG 327  
 444 CACCTGAAGCGCTGGGTGAAGGTGATCGAGGAGAAAGCTTCAAGCCCGAGGTATCC 503  
 328 CATCGCTTGTGTCCGAGAGCGTGGCTCGACGCCGACCGCTGAAGTGTGTCTCA 387  
 504 CATGTTACCGCTTGAGCGAGGCGCCACCCCCAGAACTGAAACATGTTGAAC 563  
 388 CATCGAGCAGAGTCCCGGACATCGCGAGGGGGTGCACGGCACTTCAACCAAGCGCC 447  
 564 CTTGGGCGGCGACAGGCGCGCATGCAATGTGAAAGACACATCAAGAGAGCGCG 623  
 448 CGAGGAGATCGGCGCGCGGACCAAGGCGCATGTTGGCTAGCCAGCAGAGACCCC 507  
 624 CGAGTGGAGCCGGGTGACCCCGTGCACCGCGCCCATGCGCCCGGCGCAAGTGGCGCA 683  
 508 CGAGCTGATGCCCTCAGCGAGTCTCTCGGCAACCAAGCTCGGGGCGCGCTCACCGAGCT 567

ORyza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 824)  
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 Large-scale identification of ESTs involved in the interaction  
 between rice and Magnaporthe grisea

TITLE Unpublished (2003)

JOURNAL Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu

COMMENT PCR Primers  
 FORWARD: gta aaa cga cgg cca gtc  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 12 row: P column: 19  
 Seq primer: gta aaa cga cgg cca gtc.  
 Location/Qualifiers  
 1..824  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNBd12P19"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNBd"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
 XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"

ORIGIN

Query Match 5.3%; Score 78; DB 6; Length 824;  
 Best Local Similarity 44.4%; Pred. No. 0.00032;  
 Matches 312; Conservative 0; Mismatches 390; Indels 0; Gaps 0;

204 GCAGACCGGCGAGGAGCTGAAGAGCTGTTCAACACCGTGGCCACCTGATCGCT 263  
 98 CGAGATGAGAAATGGCGCGGAGACGTTCTTCACTCGAGTCCGTGAACGAGGTCA 157

QY 264 GCACGAGAGATCGAGTCCGGACACCAAGGAGCCCTGACAGATCGAGAGAGCA 323  
DB 158 CCGGACAGAGCTGTGCGACAGGTGTGGACCGGCTGCTGACGGGTGCTTCCCGCAGGA 217  
QY 324 GAACAAGTCCGACGAGAGATCCAGAGCCGAGGCGCCGACGAGGCGAGGTAGCA 383  
DB 218 CCGGACAGAGAGTGGCGTGGACGTCGACCAAGACCAACATGTGTGTGTGG 277  
QY 384 GAACATCCCATCTGTGACAGAACTTGAAGGCGACAGTGTGACCGAGCCATAGCCCGC 443  
DB 278 CGAGATCAACCAAGGCGACCGTGAATCAAGAGATGTCTCCGACACTGCGCGG 337  
QY 444 CACCTGAACGCTGGGTGAAGTGTAGGAGAGAGGCTTCAGCCCGAGGTATGCC 503  
DB 338 CATGGCTTGTGTGCGACAGCTGTGGCTTCAAGCCGACCGCTGCAAGTGTCTGTCA 397  
QY 504 CATGTTCACCGCTTGAAGGAGGCGCCACCCCGACGACCTTGAACAGATTTGAAC 563  
DB 398 CATCGAGAGAGTGGCGCCGACATGCGCGAGGGGGTGCACGCGCACTTCAACAGCGCC 457  
QY 564 CGTGGGCGGCGACGAGGCGCCATGACATGCTGAAGACCACTCAACGAGAGCGCGC 623  
DB 458 CGAGAGATCGGCGCGCGGACGAGGCGCAATGTTGCGCTACGCGACGAGAGACCC 517  
QY 624 CGAGTGGAGCGCGTGCACCCCGTGCACGCGCGCCCATCGCCCGCGCAGATGCGGA 683  
DB 518 CGAGCTGATGCTCTTCAAGCAAGTCTCTGCGCACCAAGCTGCGCGCTTCAAGAGT 577  
QY 684 GCGCGCGGCGAGCATTCGCGGACACCAAGACCTTGCAGAGAGAGTGCCTGAT 743  
DB 578 CCGCAAGAGAGGACCTTGGCTGCTGAGCGCCGACGAGAGAGAGTCAAGTGA 637  
QY 744 GACCGAGAACCCCTTCCCGTGGGAGCATCTTCAAGCGGTGATATCTTGGCT 803  
DB 638 GTACTTCAACGACGCGCGCGCATGATGCTCCGCTGCTCAACCGTCTTATCTTCA 697  
QY 804 GAACAAGATCTGCGGATGTACAGCCCGTGCAGCTCTTCAAGAGAGGCGCCAA 863  
DB 698 CAGAGACGAGAGACCGTCAACAGAGAGATGCGCGCGACCTTCAAGAGAGATCAT 757  
QY 864 GGAGCCCTTCCGCACTACGTGAGCGCTTCTTCAAGACCTT 905  
DB 758 CAGCGCGGTATCCCGCAAGTACTTCAAGAGAGACAT 799

RESULT 10  
CB645622 851 bp mRNA linear EST 08-APR-2003  
LOCUS OSJNB07H03.f OSJNB Oryza sativa (japonica cultivar-group) cDNA  
DEFINITION clone OSJNB07H03 5', mRNA sequence.  
ACCESSION CB645622  
VERSION CB645622.1 GI:29640613  
KEYWORDS EST.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehretiaceae; Oryzaceae; Oryza.  
REFERENCE 1 (bases 1 to 851)  
AUTHORS Jantaasriyarak,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.  
TITLE Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
JOURNAL Unpublished (2003)  
COMMENT Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers

FORWARD: gta aaa cga cgg cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 07 row: H column: 03  
Seq primer: gta aaa cga cgg cca gtc.  
Location/Qualifiers  
1..851  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:3947"  
/clone="OSJNB07H03"  
/tissue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="MD10B"  
/clone\_lib="OSJNB"  
/note="Vector: plasmidscript II KS +; Site 1: EcoRI, Site 2:  
XhoI; 24 hrs after inoculation with Rice Blast (Che  
86061)"

## ORIGIN

Query Match 5.3%; Score 78; DB 6; Length 851;  
Best Local Similarity 44.4%; Pred. No. 0.00032;  
Matches 312; Conservative 0; Mismatches 390; Indels 0; Gaps 0;  
QY 204 GCAGACCGGCGAGAGAGCTGAAGAGCTTTCAACACCGTGGCCACCTGTACTGGT 263  
DB 105 GCAATTAAGAAATGCGCGGAGAGAGCTTCTTCACTTCCAGTCCGTGAACAGAGGTCA 164  
QY 264 GCACGAGAGATCGAGTTCGCGACACCAAGAGAGCCCTTGACAAATCGAGAGAGCA 323  
DB 165 CCGGACAGAGTGTGCGACCAAGTGTGAGAGCGGTGCTGACCGCTGCTGCCCGCAGGA 224  
QY 324 GAACAAGTCCGAGAGAGATCCAGAGGCGCGAGGCGCGACAGGCGCAAGGAGAGCA 383  
DB 225 CCGGACAGAGAGTGGGTGCGAGAGCTGACCAAGACCAAGATGATGATGTTGCG 284  
QY 384 GAATTAACCCATCTGTGAGAACTTGCAGAGGCGAGATGATGACCAAGGCGCATCAGCCCG 443  
DB 285 CGAGATCAACCAAGGCGACCGTGCATACGAGAAATGTCTCGCGACACTTCCCGG 344  
QY 444 CACCTGAACGCTGGGTGAAGTGTATGAGAGAGAGGCTTCAAGCCCGAGTATCC 503  
DB 345 CATGGCTTGTGTTCGCGACGAGTTCGCGCTGACGCGGACCGTGCAGAGTGTCTTCA 404  
QY 504 CATGTACCGCCCTGAGCGAGGCGCCACCCCGACAGACCTTGAACAGATTTGAAC 563  
DB 405 CATGAGCAGAGATGCGCGCCGACATGCGCAGAGGGGTGACGGCACTTCAACAGCGCC 464  
QY 564 CGTGGGCGGCGACAGGCGCGCATGACAGTCTGAAGACATCAAGAGAGAGCGCGC 623  
DB 465 CGAGAGATGCGCGCGCGCGACAGGCGCATATGTTGGGTAGCGCACCGAGAGAGCC 524  
QY 624 CGAGTGGAGCGCGTGCACCCGTGCAGCGCGGCCCATATGCCCCGCGCAGATGCGGA 683  
DB 525 CGACTGTATGCGCTTCAAGCCATGCTTGCACCAAGGCTCGCGCGCTTCAAGAGT 584  
QY 684 GCGCGCGGCGAGCATTCGCGCGGACCAACAGACCTTGAAGAGAGATGCGTGGAT 743  
DB 585 CCGGAGAGAGGCGACCTGCGCGCTTGAAGCGCGAGGAGACCAAGGTCACGTTGA 644  
QY 744 GACGAGCAACCCCGCATTCCTGGGCGACATTTAAGCGGTGATCATCTTGGGCT 803  
DB 645 GTACTTCAACGACCGCGCGCATATGTTCCCGTCCGCTGACACCGTCTCATCTCAC 704  
QY 804 GAACAAGTCTGCGATGTACAGCCCGGTGAGATCTTGAACATCAAGAGAGGCGCCAA 863  
DB 705 CAGCAGACGAGAGCGGTCAACAGAGAGATGCGCGGACCTTCAAGAGAGAGTCA 764  
QY 864 GGAGCCCTTCCGCACTACGTGAGCGCTTCTTCAAGACCTT 905  
DB 765 CAGCGCGGTATCCCGCAAGTACTTCAAGAGAGAGACAT 806

RESULT 11  
 CB673355 807 bp mRNA 1linear EST 09-APR-2003  
 LOCUS CB673355  
 DEFINITION OSJUNE07N10.f OSJUNE Oryza sativa (japonica cultivar-group) cDNA  
 accession CB673355  
 version CB673355.1 GI:29677080  
 keywords EST.  
 source Oryza sativa (japonica cultivar-group)  
 organism Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Eriactoidae; Oryzaceae; Oryza.  
 reference 1 (bases 1 to 807)  
 Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,  
 Kudrna, D., Dean, R., Soderlund, C., Wang, R. and Wang, G.  
 Large-scale identification of ESTs involved in the interaction  
 between rice and Magnaporthe grisea  
 Unpublished (2003)  
 contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR primers  
 FORWARD: gta aac cga cgg cca gtc  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 07 row: N column: 10  
 Seq primer: gta aac cga cgg cca gtc.  
 location/Qualifiers  
 1..807  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJUNE07N10"  
 /issue\_type="leaf"  
 /dev\_stage="3 week"  
 /lab\_host="PH10B"  
 /clone\_1ib="OSJUNE"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
 KhoI; 24 hrs after inoculation with Rice Blast (70-15)"  
 ORIGIN  
 Query Match 5.2%; Score 76.6; DB 6; Length 807;  
 Best Local Similarity 44.3%; Pred. No. 0.00058;  
 Matches 310; Conservative 0; Mismatches 389; Indels 0; Gaps 0;  
 Oy 204 GCAGACCGGACGAGGAGCTGAAAGGCTGTTCAACCGTGGCCACCTGTACTGCGT 263  
 Db 109 GCAGATAGAAATGGCGCGGAGAGCTTCTCTTCACTCCGAGTCCGTGAACGAGGTCA 168  
 Oy 264 GCAGGAGAGATCGAGTCCGCGACACCAAGAGGCGCTGCAACATCGAGAGAGCA 323  
 Db 169 CCCGACACAGCTGTGACACAGTGTGACGCGGCTGCTGACGCGCTGCTGCCACAGA 228  
 Oy 334 GAACAAGTCCAGCAGAAAGATCCAGCAGGCGCGCGGACCAAGGAGCAAGGTGAGCA 383  
 Db 229 CCCGACAGCAAGTGGCGTGGAGACGTGACCAAGACCAACATGATGTGTGTTGG 288  
 Oy 384 GAACATACCCATGTGTCAGAACTGCAAGGCGCCAGATGTGTCACAGGCCATAGCCCCG 443  
 Db 289 CGAGATACACCAAGGCCACCGTCATCACTAAGGAATGTCCTCGACACCTGCGCCG 348  
 Oy 444 CACCCGAGAGCGCTGGGTGAAGTGTATCGAGAGAGAGGCTTCAGCCCGAGGTATGCC 503  
 Db 349 CATGGCTTGTGTCCGACGACGTGCGCTCGAGCGCCGACCGTGAAGGTGTCTCTCA 408  
 Oy 504 CATGTTACCGCGCTGAGGAGGCGCAACCCCGACGACCTGAACATGATTTGAACAC 563

Db 409 CATCGACGACGATGCGCCCGACATCGCGAGGGGGGTGACGCGGACCTTCACCAAGCGCCC 468  
 Oy 564 CGTGGCGCGGACACGAGCGCGCCATGATGCTGAAGGACACCATCAACAGAGGCGCG 623  
 Db 469 CGAGAGATGCGCGCGCGGACACAGGCGCACTGTTGCGCTACGCCACGAGAGACCCC 528  
 Oy 624 CGAGTGGGACCGCGTGCACCCCGTGCAGCGCGGCCCATTCGCCCCCGGACATGGCGCA 683  
 Db 529 CGAGCTGATGCGCTTCAGCAGCATGTCCTGCGCCACCAAGGCTCGCGCGCCGCTCACGAGGT 588  
 Oy 684 GCCCGCGGACGACATGCGCGGACCAACAGACCTTCGACGAGGACGATGCGCTGAT 743  
 Db 589 CCGCAAGAACGCGACCTGCGCTGCTGCTGAGCGCCACGCGCAAGACCGTACCGTTGA 648  
 Oy 744 GACGCAACACCCCGCATCCCGTGGGGAGACTTACAAGGGGTGATGATCGTGGGCGCT 803  
 Db 649 GAACTTCAACGAGCGCGGCGGATGATGCTCCGCTCGCTCCACACCGTCTATCTCCAC 708  
 Oy 804 GAACAAGATCGTGCAGATGTCAGCCCGTGCATCTCTGACATCAAGAGGCGCCCA 863  
 Db 709 CAGACACGACGAGCGCGTACCAACGAGATGCGCGCGGACCTCAAGAGACAGTAT 768  
 Oy 864 GAGCGCTTCCCGGACATACGTGACCGCTTCTTCAAGAC 902  
 Db 769 CAAGCGGTATCCCGGACAGTACTGACGAGAAAG 807  
 RESULT 12  
 CK163513 1138 bp mRNA 1linear EST 05-DEC-2003  
 LOCUS CK163513  
 DEFINITION FGAS016142 Triticum aestivum FGAS. Library 4 Gate 8 Triticum  
 accession CK163513  
 version CK163513.1 GI:38993817  
 keywords EST.  
 source Triticum aestivum (bread wheat)  
 organism Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 reference 1 (bases 1 to 1138)  
 Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,  
 Genswein, B., Graf, R., Gullick, P., Hrycan, L.D., Laroche, A.,  
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,  
 Peniket, C., Roach, J.L. and Sarhan, F.  
 Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
 Unpublished (2003)  
 contact: Wm L Crosby  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science  
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033  
 Email: fgas.est@cs.usask.ca  
 This sequence is the direct result of the base calling software  
 phred (default parameters). It is the raw base calls. To aid in the  
 identification of the high quality insert the software Lucy  
 (default parameters) has been run on this sequence. Lucy identified  
 the region [8,827].  
 Plate: LAB008 row: F column: 04.  
 location/Qualifiers  
 1..1138  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4565"  
 /note="Organ: Triticum aestivum FGAS; Library 4 Gate 8"  
 /clone\_1ib="Triticum aestivum FGAS; Library 4 Gate 8"  
 Conditions for growth: Seeds were germinated in a  
 water-saturated mix (50% black earth and 50% ProMix) in a  
 growth chamber for 7 days under an irradiance of 200 mmol  
 m-2 sec-1. The temperature was maintained at 20 degrees C  
 with a 15-hr photoperiod under a relative humidity of 70%.

After this period watering of plants was stopped. Four time points were sampled during a two week period; the first after wilting was observed and the last, two weeks later, consisted of live crown and leaf tissue (leaf tissue that was yellow was not included in sampled material). First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

## ORIGIN

Query Match 5.2%; Score 76.6; DB 7; Length 1138;  
Best Local Similarity 49.0%; Pred. No. 0.00059;  
Matches 291; Conservative 0; Mismatches 295; Indels 8; Gaps 3;

124 GAGAAATTGCGCTTGAACCCCGCTGTTGAGACCAAGGAGGCTGCAAGATCATC 183  
123 GAGATACCGCCACCAACCACTGAACTGATGACACCTCCGCGACGAGCGCAATTC 291  
184 GCGCAGCTGCAACCCCGCTGAGACCGCGCAAGAGCTGAAGCTGTTCAACCC 243  
183 CTCAACATGCTCTCAAGCTCATCGCGCGCAAGAGACCATGAGTCCGCTTCAACC 351  
224 GTGGCCACCTGTACTGCTGTGCAAGAAATGAGTCCGCGACCAAGAGGCGCTG 303  
223 GCGTACTCTCTCTGCTGCGCACCGCGCTGCGCATCCCGACGAGCGACCATTTGGCCATG 411  
304 GCAAGANTGAGAGAGAGAGAAACAAGTCCAGCAGAAAGATTCAGAGGCGCGCCG 363  
412 GATCTCAACCGGAGAACTACAGAGCTGGGGCTGCGGTGATGAGAGGCGCGCTGGCG 471  
364 GACAGGGGAGAGTGAAGCAAGAACTACCCCATGCTGAGAACTGAGGCGCGAGTGGTG 423  
472 CACAAAGTACGCTTCCGCGAG--GGCCCGCGCGCTGCCCTGAGCGCTGGAG 528  
424 CACAGAGCCATAGCCCGCGCAACCTGAAAGCTGGGTGAAGTATGAGAGAGAGCC 483  
529 GACGAGGCGCAACCGAGGACCTTCACTTCTGCTTGTGAGCGCGCAAGAGAACTAC 588  
484 TTGAGCGCGAGGTATCCCATGTTTCAACCGCTGAGGAGGCGCGCAACCCCGAGAC 543  
589 CTCACATACCAAGAGGCGCTCATGAAAGCTGCTCAAGTGGCGGCGCTCTCGGCTACGAC 648  
544 CTGAACAGATGTTGAACAACCGTGGCGCGCGCAACGAGCGCGCATGAGATGTTGAAGAC 603  
649 ---AACACCTCTGAAAGGCTCCGCTGCTGCTCCCGCGAGCGCCCGCAAGATAC 705  
604 ACCATCAACGAGAGGCGCGCGAGTGGAGCCGC--GTGCAACCCCGTGAACGCGCGCGCA 661  
706 ATCCGCTACTACCGGAGCTTGTCTCGAGCTCAACAGGCGCTCGCGCGAGCAAGCGCG 765  
QY 662 TGGCCCGCGCGAGATGCGCGAGCGCGCGCGAGGAGCATGCGCGGACCAACA 715  
DB 766 TCGAGATCTGCGAGCTCCCGCTGCGGAGGCGCATCACTCTGCGCGCGCGCA 819

RESULT 13  
CB681973 812 bp mRNA linear EST 09-APR-2003  
LOCUS OSUNEf08M09.f OSUNEf Oryza sativa (japonica cultivar-group) cDNA  
DEFINITION clone OSUNEf08M09 5', mRNA sequence.  
ACCESSION CB681973  
VERSION CB681973.1 GI:29686698  
KEYWORDS EST.  
ORGANISM Oryza sativa (japonica cultivar-group)  
SOURCE Oryza sativa (japonica cultivar-group)  
Bukiryocta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Bhurarioidae; Oryzaceae; Oryza.  
REFERENCE 1 (bases 1 to 812)  
AUTHORS Jantaauriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R., and Wang,G.  
TITLE Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea

## JOURNAL

Unpublished (2003)

## COMMENT

Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aac cga cga cga gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 08 row: M column: 09  
Seq primer: gta aac cga cga cga gtc.  
Location/Qualifiers

## FEATURES

## Source

1. 812  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSUNEf08M09"  
/tissue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/note="Vector: pBlueScript II KS +; Site\_1: EcoRI; Site\_2:  
XhoI; Uninfected Control"

## ORIGIN

Query Match 5.2%; Score 76.4; DB 6; Length 812;  
Best Local Similarity 44.3%; Pred. No. 0.00064;  
Matches 311; Conservative 0; Mismatches 391; Indels 0; Gaps 0;

QY 204 GACAGCGGCAAGAGAGTGAAGCTGTTCAACACCGTGCCACCTGTACTGGCT 263  
DB 105 GCAATGAAATGAGGCGGAGAGAGCTTCTTCACTCCGAGTCCGTAACGAGGCTCA 164  
QY 264 GACAGAAAGTCAAGTTCGCGGACACCAAGAGGCGCTTGAACAAGATGAGAGAGCA 323  
DB 165 CCGGACAAAGGTGTGCGACAGGTGTGAGAGCGGTGTCTGACGCGTCCGCGCGA 224  
QY 324 GAACAAGTCCAGCAAGATTCAGAGGCGCGAGCGCGCGCAAGGCGAGGTGAGCA 383  
DB 225 CCGGACAGCAAGGTGTGCGAGAGCTGACCAAGACCAACATGATGATGTTCG 284  
QY 384 GAATCAACCATGCTGCAAGAACTGCAAGGCGCAAGTGTGCAACAGCCATCAGCCCG 443  
DB 285 CGAGTCAACCAAGGCGCACTGCTGACTACGAAAGTGTGCGCGCACCTCGCGCG 344  
QY 444 CACCTGAACCGCTGTGAGAGTGTGAGAGAGAGAGGCTTCAAGCCCGGAGTATCC 503  
DB 345 CATGCGCTTGTGTGTCGAGAGTGTGAGCGCTGAGCGCGCAAGGTGTGTTAA 404  
QY 504 CATGTTCAACCGCTGAGAGGCGCGCAACCCCGAGAGCTGAACAGATGTTGAAC 563  
DB 405 CATGAGAGAGAGTGTGCGGACATCGCGCAAGGAGGTGACGCGCATTCACCAAGCGCC 464  
QY 564 CTGAGCGGCGCAACAGCGCGCATGACAGATCTGAAGAGCAACATCAAGAGAGCGCC 623  
DB 465 CGAGGAGATCGCGCGCGCGAGCAAGGCGCATGTTGGCTACCGCACCGAGAGCC 524  
QY 624 CGATGAGAGCGCGTGTGACCCCGTGTGACGCGCGCGCATTCGCGCGCGAGTCCGGA 683  
DB 525 CGAGTGTATGCGCTTCAAGCAAGTCTTCCGACCAAGCTTGGCGCGCGCTTCAAGAGT 584  
QY 684 GCGCGCGCGAGAGCATGCGCGGACCAACAGAGACCTTGAAGAGAGATCGCTTGAT 743  
DB 585 CCGAAGAGCGAGCATGCGCGCTGAGCTGAGCGCGAGAGAGAGAGAGAGAGAGAGTGA 644  
QY 744 GACAGCAACCCCGCATCCCGTGGGAGAGATTAACAGGCGGTGATCATCTGGGCT 803  
DB 645 GTACTGAAAGAGCGCGCGCATGATGTCCTCGCGGTGACACCGGCTCATCTTCAC 704



QY 804 GAACAAGATCGTCGGATGTACAGCCCCCTGAGCATCTGATCATCAAGAGGGCCCA 863  
| | | | |  
DB 705 CCAGCAGCAGACCGTCAACAGCAGAGATCGCCCGGACCTCAAGAGACAGCTCAT 764  
| | | | |  
QY 864 GGAAGCCCTTCGGGACTACGTGACCGCTTCTTCAAGACCT 905  
| | | | |  
DB 765 CAAGCCGGTCATTCGCGACAAAGTACCTCGACGAGAAACCAT 806  
| | | | |  
  
RESULT 14  
CB684834 832 bp mRNA linear EST 09-APR-2003  
LOCUS OSJNEF1424.f OSJNEF Oryza sativa (japonica cultivar-group) cDNA  
DEFINITION clone OSJNEF1424 5', mRNA sequence.  
ACCESSION CB684834  
VERSION CB684834.1 GI:29688559  
KEYWORDS EST.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 832)  
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
Unpublished (2003)  
JOURNAL Contact: Rod Wing  
COMMENT Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aac cga cgg cca gtc g  
BACKWARD: gga aac agc tat gac cat g  
Plate: 14 row: 1 column: 24  
Seq primer: gta aac cga cgg cca gtc g.  
Location/Qualifiers  
1..832  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSJNEF1424"  
/cissue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OSJNEF"  
/note="Vector: pBluescript II KS +; Site\_1: EcoRI; Site\_2:  
XhoI; Uninfected Control"

DB 277 CGAGATCAACCAAGCCACCGCTGATCAAGAAATCTGTCGGGACACCTCGCCGG 336  
| | | | |  
QY 444 CACCTGAAGCCTTGAGTGAAGTGTAGAGAAAGCCTTCAAGCCCGAGGTATCC 503  
| | | | |  
DB 337 CATCGGCTTGTGTCCGACGACGTGCGCTCGACCGCCGACCGCTGCAAGGTCTGTCAA 396  
| | | | |  
QY 504 CATGTTCAACCGCCCTGAGGAGGGGCGCCACCCCGACGACTGAACATGTTGAAC 563  
| | | | |  
DB 397 CATCGACGACAGTGCCTCCGACATGCGCAGAGGGGTGCGGCCACTTCAACAGCGCC 456  
| | | | |  
QY 564 CGTGGCGGCGCCACAGCGCCGATGATGCTGAAGGACACCATCAAGAGAGGCGCC 623  
| | | | |  
DB 457 CGAGAGATGCGGCGCGCGGACCAAGGCCACATGTTGCGCTACGCCAGAGAGACCC 516  
| | | | |  
QY 624 CGAGTGGACCGCGTGCACCTCGTGACCGCGCCCATATGCCCCCGGCGAGATGCGCA 683  
| | | | |  
DB 517 CGAGCTGATGCCCCCTCAGCCAGCTCTCGCCACCAAGCTCGGCGCGCTCACAGGT 576  
| | | | |  
QY 684 GCGCGCGGCGACGATGCGCGGACCAACGACCTCGAGAGAGATGCGCTGAT 743  
| | | | |  
DB 577 CCGAAGAACGGACACTGCGCTGCTGCTCAGGCCCAAGCGCAAGACCAAGTCACTTGA 636  
| | | | |  
QY 744 GACCAAGACCCCGCATCCCGTGGGCGACATCTTAAGCGGTGATCATCTGGAGCT 803  
| | | | |  
DB 637 GTACTTCAACGACGCGCGGCGCATGTCTCCCGCTGCACACCGTCTCATCTCCAC 696  
| | | | |  
QY 804 GAACAAGATCGTCGGATGTACAGCCCCCTGAGCATCTGACATCAAGAGGGCCCA 863  
| | | | |  
DB 697 CCAGCAGCAGACCGTCAACCAAGATGCGCGCGGACCTCAAGAGACAGTCA 756  
| | | | |  
QY 864 GGAAGCCCTTCGGGACTACGTGACCGCTTCTTCAAGACCT 905  
| | | | |  
DB 757 CAAGCCGGTCATTCGCGACAAAGTACCTCGACGAGAAACCAT 798  
| | | | |  
  
RESULT 15  
BF259495 759 bp mRNA linear EST 22-OCT-2001  
LOCUS HVSMF0019H19f Hordeum vulgare seedling root EST library HVCNA0007  
DEFINITION (Etisolated and unstressed) Hordeum vulgare subsp. vulgare cDNA  
clone HVSMF0019H19f, mRNA sequence.  
ACCESSION BF259495  
VERSION BF259495.2 GI:11120022  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
1 (bases 1 to 759)  
Wing,R., Close,T.J., Kleinhofe,A., Wise,R., Begum,D., Fritsch,D.,  
Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W.,  
Fenton,R.D., Oates,R. and Main,D.  
Development of a genetically and physically anchored EST resource  
for barley genomics: Morex unstressed seedling root cDNA library  
Unpublished (2001)  
On Nov 16, 2000 this sequence version replaced gi:11188608.  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total hg bases = 388  
Seq primer: AATTAACCTCACTAAGG  
High quality sequence stop: 758.  
Location/Qualifiers  
1..759  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Morex"

```
/sub_species="vulgare"  
/db_xref="taxon:112509"  
/clone="HVSMEF0019H19F"  
/tissue_type="Seedling root"  
/lab_host="HJC121"  
/clone_lib="Hordeum vulgare seedling root EST library  
HVCDA0007 (Etolated and unstressed)"  
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
Seeds were surface sterilized then germinated under axenic  
conditions in the dark at room temperature on filter paper  
with water, nystatin and cefotaxime in covered  
crystallization dishes. Five-day old seedling roots were  
then harvested, total RNA was prepared, poly(A) RNA was  
purified, one primary unamplified cDNA library was made,  
and 1 million pfu were in vivo excised to give Bluescript  
SK(-) cDNA phagemids. These steps were performed in the TU  
close laboratory at the University of California,  
Riverside (Choi, Close, Fenton). Phagemids were plated and  
picked at the Clemson University Genomics Institute (CUGI)  
(Begum, Palmer, Fritsch, Atkins and Wing). Plasmid DNA  
preparations, DNA sequencing and sequence analysis were  
performed at CUGI (Wing, Yu, Fritsch, Henry, Simmons,  
Oates, Rambo, Main). The sequence has been trimmed to  
remove vector sequence and contains a minimum of 100 bases  
of phred value 20 or above. For more details on library  
preparation and sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders Also  
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genome. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html)"
```

## ORIGIN

Query Match 5.2%; Score 76.2; DB 2; Length 759;

Best Local Similarity 47.7%; Pred. No. 0.00069;

Matches 284; Conservative 0; Mismatches 305; Indels 6; Gaps 2;

```
QY 124 GAGAAATTGCGCCCTGAAACCCCGGCTGTGAGAGCCAGGAGGCTGCAAGCATATC 183  
DB 13 GAGATCACCGCCAAACCCATGAACTGATGACGACGTCGGCGAGCGAGGCGCAGTTC 72  
QY 184 CGCAGCTGACACCCCGCCCTGACAGACCGGACGAGAGAGCTGAAGGCTTTCAACACC 243  
DB 73 CTCAAATGCTGTCTCAAGCTCATCGGCGCCAAAGAACATGAGATCGGCGCTACACG 132  
QY 244 GTGGCCACCTGTATGCTGTCACGAAAGATGAGGTCCGCGACACCAAGAGGCGCTG 303  
DB 133 GGCTACTCTCTGTGTCGCAACCGGCTCGGCATCCCGACGAGCGACATCTTGGCCATG 192  
QY 304 GACAGATGAGAGGAGAGCAAAACAAGTGCAGACAGAAATCCAGAGGCCAGGCCGCC 363  
DB 193 GACATCAACCGCAGAACTACGAGCTGGGGCTGCTGTGATCGAAGAGCCGCGTGCGG 252  
QY 364 GACAAAGGCAAGGTGAGCCAACTACCCCATGTGACAGACTTGCAAGGCGCATGTG 423  
DB 253 CACAAATGATCTTCGCGAGAGGCCCGGGGCTC---CCGTCCTTGAGCGCCTCTCGAG 309  
QY 424 CACCAGGCGATCAGCCCGCGACCTGAAAGCCTGGGTGAAGGTATCGAAGAAAGGCC 483  
DB 310 GACGAGGCCAAACACGCGACCTTCGACTTGTCTTGTTGAGACGCGCAAGGACAACTAC 369  
QY 484 TTGAGCCCGAGGTGATCCCATGTTCAACGCGCTGAGAGGAGCGCCACCCCGAGAC 543  
DB 370 CTCAACTACCAACGAGGCGCTCATGAAGCTGTCAAGGTGCGCGGCTCTCTGGCTACGAC 429  
QY 544 CTGAACAGATGTTGAACACCGTGGGGGCGCACAGGCGCCATGCAATGCTGAAGAC 603  
DB 430 ---AACAACCTCTGGAACGCTCCGTCTGCTCCCGCGAGCGCCCATGCGCAAGTAC 486  
QY 604 ACCATCAACGAGAGGCGCGGAGTGGGACCGGCTGACCCCGTGACCGCGGCGCCATC 663  
DB 487 ATCGGCTACTACGCGCATCTGTCTGCACTCAACAAGGCCCTCGCGCGGACGAGCGC 546
```

```
QY 664 GCCCGCGCGATGCGCGAGACCCCGGCGAGCGACATCGCCGCGACCAACAGCA 718  
DB 547 GTGAGATCTGCGCAGCTCCCGTGGCGAGCGGATCACCTCTGCGCGCGCGCA 601
```

Search completed: March 13, 2005, 00:13:44  
Job time : 5214.18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 12, 2005, 16:41:47 ; Search time 261.349 Seconds

(without alignments)  
925.839 Million cell updates/sec

Title: US-09-475-704A-3

Perfect score: 1479

Sequence: 1 atggggccggccgagccagcat.....accgacccttgagccagtaa 1479

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405566

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA:\*

1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*\n2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*\n3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*\n4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*\n5: /cgn2\_6/prodata/1/ina/PTDUS\_COMB.seq:\*\n6: /cgn2\_6/prodata/1/ina/backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1197.8	81.0	1515	4 US-09-475-515-4	Sequence 4, Appl1
2	1197.8	81.0	4472	4 US-09-475-515-75	Sequence 75, Appl1
3	1197.8	81.0	4608	4 US-09-475-515-76	Sequence 76, Appl1
4	1197.8	81.0	4689	4 US-09-475-515-74	Sequence 74, Appl1
5	1197.8	81.0	4766	4 US-09-475-515-73	Sequence 73, Appl1
6	1195.8	80.9	2031	4 US-09-475-515-7	Sequence 7, Appl1
7	1114.6	75.4	1853	4 US-09-475-515-5	Sequence 5, Appl1
8	1114.6	75.4	1865	4 US-09-475-515-78	Sequence 78, Appl1
9	1114.6	75.4	1865	4 US-09-475-515-79	Sequence 79, Appl1
10	1114.6	75.4	4319	4 US-09-475-515-6	Sequence 6, Appl1
11	1025.2	69.3	1268	4 US-09-475-515-9	Sequence 9, Appl1
12	976.6	66.0	1521	4 US-09-952-060-27	Sequence 27, Appl1
13	976.6	66.0	1532	4 US-09-974-702-1	Sequence 1, Appl1
14	976.6	66.0	1532	4 US-09-918-443-1	Sequence 1, Appl1
15	976.6	66.0	37474	4 US-09-952-060-25	Sequence 25, Appl1
16	973.6	65.8	4053	4 US-09-952-060-34	Sequence 34, Appl1
17	949.4	64.2	4307	4 US-09-936-572-2	Sequence 2, Appl1
18	947.6	64.1	4307	4 US-09-552-950-2	Sequence 2, Appl1
19	939.8	63.5	4327	4 US-09-936-572-14	Sequence 14, Appl1
20	939.8	63.5	4353	4 US-09-936-572-13	Sequence 13, Appl1
21	939.8	63.5	4642	4 US-09-936-572-12	Sequence 12, Appl1
22	939.8	63.5	9772	4 US-09-552-950-5	Sequence 5, Appl1
23	933	63.1	1482	4 US-09-818-443-4	Sequence 4, Appl1
24	930	60.3	1479	4 US-09-952-060-32	Sequence 32, Appl1
25	744.6	50.3	8366	4 US-09-872-733A-6	Sequence 6, Appl1
26	738.2	49.9	1476	4 US-09-184-418C-100	Sequence 100, Appl1
27	738.2	49.9	8959	4 US-09-184-418C-11	Sequence 11, Appl1

28	733.4	49.6	4338	4 US-09-872-733A-1	Sequence 1, Appl1
29	731.4	49.5	9913	4 US-09-827-688-11	Sequence 11, Appl1
30	729.8	49.3	1485	4 US-09-184-418C-73	Sequence 73, Appl1
31	729.8	49.3	9010	4 US-09-184-418C-8	Sequence 8, Appl1
32	728.4	49.2	12523	4 US-09-991-258-1	Sequence 1, Appl1
33	726	49.1	1476	4 US-09-991-258-4	Sequence 4, Appl1
34	710.8	48.1	1486	4 US-09-184-418C-38	Sequence 38, Appl1
35	710.8	48.1	8992	4 US-09-184-418C-4	Sequence 4, Appl1
36	701.2	47.4	1496	4 US-09-184-418C-82	Sequence 82, Appl1
37	701.2	47.4	8972	4 US-09-184-418C-9	Sequence 9, Appl1
38	659.8	44.6	1503	4 US-09-184-418C-29	Sequence 29, Appl1
39	659.8	44.6	8953	4 US-09-184-418C-3	Sequence 3, Appl1
40	658.8	44.5	1479	4 US-09-184-418C-12	Sequence 12, Appl1
41	658.8	44.5	8968	4 US-09-184-418C-1	Sequence 1, Appl1
42	656.6	44.4	4307	4 US-09-184-418C-48	Sequence 48, Appl1
43	656.6	44.4	8966	4 US-09-184-418C-5	Sequence 5, Appl1
44	652.6	44.1	1494	4 US-09-184-418C-47	Sequence 47, Appl1
45	650.2	44.0	4307	4 US-09-552-950-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-09-475-515-4 ; Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1515  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; US-09-475-515-4

Query Match 81.0%; Score 1197.8; DB 4; Length 1515;  
Best Local Similarity 89.3%; Pred. No. 9.3e-177;  
Matches 1347; Conservative 0; Mismatches 132; Indels 30; Gaps 4;

QY	1	ATGGGCGCGCGCGCGCAATCTGCGCGCGCGCGCAAGCTGAGCCTCGGAGCGCATCCGC	60
DB	7	ATGGGCGCGCGCGCGCGCAATCTGCGCGCGCGCGCAAGCTGAGCCTCGGAGCGCATCCGC	66
QY	61	CTGGCGCGCGCGCGCGCAAGAGTCTAATGATGAAGCACTCTGATGGGCGCGCGCGAG	120
DB	67	CTGGCGCGCGCGCGCGCAAGAGTCTAATGATGAAGCACTCTGATGGGCGCGCGCGAG	126
QY	121	CTGGAAGATTGCGCCTGAAACCCCGCGCTGTGAGAGCAAGGAGCGCTGCAAGCAGATC	180
DB	127	CTGGAAGCGTTGCGCGTGAACCCCGCGCTGTGAGAGCAAGGAGCGCTGCGGCGAGTC	186
QY	181	ATCCGCGAGCTGCACCCCGCGCTGTGAGAGCGGAGGAGGAGCTGAAGAGCTGTTCAC	240
DB	187	CTGGGCGAGCTGCACCCCGCGCTGTGAGAGCGGAGGAGGAGCTGTTCAC	246



```
QY 361 GCCG-----ACAAGGCAAGTGAAGCCAGAACTACCCATCGTGCAGAACTTG 408
    |||
Db 3299 GCCGCGGCGGACCGGCAACAGCAAGCCAGGTGAGCCAGAACTACCCATCGTGCAGAACTTG 3358
QY 409 CAGGGCAGATGGTGAACCAAGGCATCAGCCCCCGGCACTGTAAAGCTTGGGTGAAGTGTG 468
    |||
Db 3359 CAGGGCAGATGGTGAACCAAGGCATCAGCCCCCGGCACTGTAAAGCTTGGGTGAAGTGTG 3418
QY 469 ATCGAGAGAAAGGCTTCAAGCCCGGAGTGAATCCCATATTTACAGCCCTTGAGCGAGGGC 528
    |||
Db 3419 GTGAGAGAAAGGCTTCAAGCCCGGAGTGAATCCCATATTTACAGCCCTTGAGCGAGGGC 3478
QY 529 GCCACCCCCCAGGACCTTGAACAAGATGTGAACACCGTGGCGGCGCAACAGGCGCCGATG 588
    |||
Db 3479 GCCACCCCCCAGGACCTTGAACAAGATGTGAACACCGTGGCGGCGCAACAGGCGCCGATG 3538
QY 589 CAGATGCTGAAGACACCATCAACGAGAGGCCCGGAGTGGAGCCGCTGACCCCGTGTG 648
    |||
Db 3539 CAGATGCTGAAGAGACCATCAACGAGAGGCCCGGAGTGGAGCCGCTGACCCCGTGTG 3598
QY 649 CACGCGGCGCCATCGCCCCCGGCGGAGTGGCGGAGCCCGGCGGCACTCGCGGCG 708
    |||
Db 3599 CACGCGGCGCCATCGCCCCCGGCGGAGTGGCGGAGCCCGGCGGCACTCGCGGCG 3658
QY 709 ACCACAGACCCCTGCAAGAGAGATGCGCTGGATGACAGCAACCCCGCATCCCGCGT 768
    |||
Db 3659 ACCACAGACCCCTGCAAGAGAGATGCGCTGGATGACAGCAACCCCGCATCCCGCGT 3718
QY 769 GCGCATCTCAAGAGCGGTGATCATCTGGGCTTGAAACAAGTCTGCGATGTACAC 828
    |||
Db 3719 GCGCATCTCAAGAGCGGTGATCATCTGGGCTTGAAACAAGTCTGCGATGTACAC 3778
QY 829 CCGGTGAGATCTTGACATCAAGCGGCCCCAAGAGCCCTTCCGCGATTAAGTGTAC 888
    |||
Db 3779 CCGGTGAGATCTTGACATCAAGCGGCCCCAAGAGCCCTTCCGCGATTAAGTGTAC 3838
QY 889 CGCTTCTTCAAGACCCCTGCGCGCGGAGAGACCCAGAGGTGAAGAACTGATGAC 948
    |||
Db 3839 CGCTTCTTCAAGACCCCTGCGCGCGGAGAGACCCAGAGGTGAAGAACTGATGAC 3898
QY 949 GACACCTGCTGTGAGAGACGCAACCCGCACTGCAAGACCATCTGCGCTCTCGGC 1008
    |||
Db 3899 GACACCTGCTGTGAGAGACGCAACCCGCACTGCAAGACCATCTGCGCTCTCGGC 3958
QY 1009 CCGGCGGCGGCTTGAAGAGATGATGACCGCTGCGAGGGGTGGGCGGCGCCAGCCAC 1068
    |||
Db 3959 CCGGCGGCGGCTTGAAGAGATGATGACCGCTGCGAGGGGTGGGCGGCGCCAGCCAC 4018
QY 1069 AAGGCGCGGTGCTGGCGAGGCGATGAGCCAGGCGCA-----CGAGCGTATGATG 1122
    |||
Db 4019 AAGGCGCGGTGCTGGCGAGGCGATGAGCCAGGCGCA-----CGAGCGTATGATG 4078
QY 1123 CAGAAAGCACTTCAAGGAGCCCGCGCGCATGTCMAAGTGTTCMAATGCGCGCAAGAG 1182
    |||
Db 4079 CAGGCGGCGCACTTCCGCAACCAAGGAGAGACCTGTAAGTGTTCMAATGCGCGCAAGAG 4138
QY 1183 GCGCACTGCGCGCACTGCGCGCGCGCGCGCGAGAGAGGCTGCTGAAGTGGCGCAAG 1242
    |||
Db 4139 GCGCACTGCGCGCACTGCGCGCGCGCGCGCGAGAGAGGCTGCTGAAGTGGCGCGCG 4198
QY 1243 GAGGGCGCAACGATGAAGAGCTGCAACCGAGCGCGAGCGCAACTTCTGGGCGAAGATG 1302
    |||
Db 4199 GAGGGCGCAACGATGAAGAGCTGCAACCGAGCGCGAGCGCAACTTCTGGGCGAAGATG 4258
QY 1303 CCAAGCCCAAGAGGCGCGCGCGCACTTCTGCAAGGCGCGCGCGAGCCACCGCGCCG 1362
    |||
Db 4259 CCAAGCCCAAGAGGCGCGCGCGCACTTCTGCAAGGCGCGCGCGAGCCACCGCGCCG 4318
QY 1363 CCGCGCGAGAGCTTCCGCTTC-----GAGAGACCAACCCCGCGCGCAAGAGAGAG--- 1413
    |||
Db 4319 CCGCGAGAGAGCTTCCGCTTCGCGAGAGAGAGACCAACCCCGCGCGCAAGAGAGAGGCC 4378
```

```
QY 1414 ---ACGAGAGACCGGAGAGCCCTGACAGCCCTGAAGAGGCTTGTGCGAGACGACCCCTTG 1470
    |||
Db 4379 ATCGACAGAGAGCTGTACCCCTTGACAGCCCTGAGCGGAGGCTGTGCGAGACGACCCGAC 4438
QY 1471 AGCCAGTAA 1479
    |||
Db 4439 AGCCAGTAA 4447

RESULT 3
US-09-475-515-76
; Sequence 76, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 76
; LENGTH: 4608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp160.modsf162.delv2.gag.modsf2
US-09-475-515-76

Query Match      81.0%; Score 1197.8; DB 4; Length 4608;
Best Local Similarity 89.3%; Pred. No. 9,3e-177;
Matches 1347; Conservative 0; Mismatches 132; Indels 30; Gaps 4;

QY 1 ATGGGCGCGCGCGCGAGATCTGCGCGGCGGCAAGCTTGGAGCGCTTGGAGCGATCCGC 60
    |||
Db 3075 ATGGGCGCGCGCGCGAGAGGTGCTGAGCGGCGGCGAGCTGAGCAAGTGGAGAGATCCGC 3134
QY 61 CTGCGCGCGCGCGGCAAGAGTCTACATGATGAAGCACTGTGTGGGCCAGCGCGAG 120
    |||
Db 3135 CTGCGCGCGCGCGGCAAGAGATGAAGATGAAGCAATGTGTGGGCCAGCGCGAG 3194
QY 121 CTGGAAGAGTGGCTTGAACCCCGGCTGCTGAGAGACCAAGAGGCGTGCAGACAGATC 180
    |||
Db 3195 CTGGAAGAGTGGCTTGAACCCCGGCTGCTGAGAGACCAAGAGGCGTGCAGACAGATC 3254
QY 181 ATCGGCGAGTGCACCCCGGCTGCAAGCCGAGCGAGAGCTGAAGAGCTGTTCAC 240
    |||
Db 3255 CTGGGCGAGCTGCACCCCGGCTGCAAGCCGAGCGAGAGAGCTGCAGCTGTACAC 3314
QY 241 ACCGTGGCACCTGTACTGTGCTGCAAGAGATCGAGGTCCGAGACCAAGAGGCC 300
    |||
Db 3315 ACCGTGGCACCTGTACTGTGCTGCAAGAGATCGAGGTTCGAGCTGTACAC 3374
QY 301 CTGGAAGAGTGAAGAGAGAGCAAGAGTGCAGCGAGAGATCAAGAGGCCAGGCGC 360
    |||
Db 3375 CTGGAAGAGTGAAGAGAGAGCAAGAGTGCAGCGAGAGAGGCCAGGCGCGCCGCG 3434
QY 361 GCCG-----ACAAGGCAAGTGAAGCCAGAACTACCCATCGTGCAGAACTTG 408
    |||
Db 3435 GCCGCGGCGACCGGCAACAGCAAGGAGTGAAGCCAGAACTACCCATCGTGCAGAACTTG 3494
QY 409 CAGGCGCAGATGGTGAACCAAGGCATCAGCCCCCGGCACTGTAAAGCTTGGGTGAAGTGTG 468
    |||
```

Db 3495 CAGGGCCAGATGTGACCAAGGCGATCAAGCCCCCGACCTTGAAAGCGCTGGTGAAGTG 3554  
QY 469 ATCGAGAGAGAGGCTTTCAGCCCCCGAGGTGATCCCATTTTCAACGGCCCTGAGCGAGGCG 528  
Db 3555 GTGAGAGAGAGGCTTTCAGCCCCCGAGGTGATCCCATTTTCAAGCGCCCTGAGCGAGGCG 3614  
QY 529 GCCACCCCCCAGAGACTGGAACAAGATGTTGAACACCGTGGGGGGGACCAAGGCGCGCATG 588  
Db 3615 GCCACCCCCCAGAGACTGGAACAAGATGTTGAACACCGTGGGGGGGACCAAGGCGCGCATG 3674  
QY 589 CAGATGCTGAAGAGACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGACACCCCGTG 648  
Db 3675 CAGATGCTGAAGAGACCATCAACGAGAGGCGCGCGAGTGGAGCCGCTGACACCCCGTG 3734  
QY 649 CAGCGCGCGCCCATGCGCCCGCCAGATGCGCGAGCCCGCGGAGCGAGCATGCGCGCG 708  
Db 3735 CAGCGCGCGCCCATGCGCCCGCCAGATGCGCGAGCCCGCGGAGCGAGCATGCGCGCG 3794  
QY 709 ACCACAGACCCCTGACAGAGAGATGCGCTGATGACAGGACCCCGCCATCCCGCGTG 768  
Db 3795 ACCACAGACCCCTGACAGAGAGATGCGCTGATGACAGGACCCCGCCATCCCGCGTG 3854  
QY 769 GCGCATCTTACAGCGGTGATCATCTGGGCTGAAACAGATGCTGCGAGTGAACG 828  
Db 3855 GCGCATCTTACAGCGGTGATCATCTGGGCTGAAACAGATGCTGCGAGTGAACG 3914  
QY 829 CCGTGTGAGATCTTGACATCAAGAGGCGCCCAAGAGGCTTTCGCGCATCTAGTGAGC 888  
Db 3915 CCGTGTGAGATCTTGACATCAAGAGGCGCCCAAGAGGCTTTCGCGCATCTAGTGAGC 3974  
QY 889 CCGTGTGAGATCTTGACATCAAGAGGCGCCCAAGAGGCTTTCGCGCATCTAGTGAGC 948  
Db 3975 CCGTGTGAGATCTTGACATCAAGAGGCGCCCAAGAGGCTTTCGCGCATCTAGTGAGC 4034  
QY 949 GACACCTGCTGTGAGAGACGCAACCCCGCATGCAAGACCATCTGCGCGCTCTGCGC 1008  
Db 4035 GACACCTGCTGTGAGAGACGCAACCCCGCATGCAAGACCATCTGCGCGCTCTGCGC 4094  
QY 1009 CCGCGGCGCGCTGTGAGAGATGATGACCGCTGCGAGGGGTGGGGCGCCCGACCGAC 1068  
Db 4095 CCGCGGCGCGCTGTGAGAGATGATGACCGCTGCGAGGGGTGGGGCGCCCGACCGAC 4154  
QY 1069 AAGGCGCGGCTGTGAGCGAGGCGAGTGAAGCGAGCGCAACA-----CCAGCTGATGATG 1122  
Db 4155 AAGGCGCGGCTGTGAGCGAGGCGAGTGAAGCGAGCGCAACA-----CCAGCTGATGATG 4214  
QY 1123 CAGAAAGCACTTCAAGGGCCCGCGCGCATCTGTAAGTGTTCATCTGCGGCAAGAG 1182  
Db 4215 CAGCGCGGCACTTCCGCAACAGCGGAGAGACCGTCAAGTGTTCATCTGCGGCAAGAG 4274  
QY 1183 GCGCATCTGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242  
Db 4275 GCGCATCTGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4334  
QY 1243 GAGGCGCAACAGATGAAGACTGCAACGAGCGCGCAAGCGCACTTCTGCGGCAAGTCTG 1302  
Db 4335 GAGGCGCAACAGATGAAGACTGCAACGAGCGCGCAAGCGCACTTCTGCGGCAAGTCTG 4394  
QY 1303 CCGACGCAAGAGGCGCGCGCGCACTTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCG 1362  
Db 4395 CCGACGCAAGAGGCGCGCGCGCGCACTTCTGCGAGAGCGCGCGCGCGCGCGCGCGCG 4454  
QY 1363 CCGCGCGAGAGTTCGCGCTT-----GAGAGATCAACCCCGCGCAAGAGAGAG--- 1413  
Db 4455 CCGCGAGAGAGTTCGCGCTT-----GAGAGATCAACCCCGCGCAAGAGAGAGAG--- 4514  
QY 1414 ---AACAGAGACCGGAGAGCTTGAACAGCTGAAGAGCTGTTGGGCAACAGACCCGCTG 1470  
Db 4515 ATCGAGAGAGAGCTGTATCCCTTGAACAGCTGCGGAGCTGTTGGGCAACAGACCCGAGC 4574  
QY 1471 AGCCAGTAA 1479  
Db 4575 AGCCAGTAA 4583

RESULT 4  
US-09-475-515-74  
Sequence 74, Application US/09475515A  
Patent No. 6602705  
GENERAL INFORMATION:  
APPLICANT: BARNETT, Susan  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: SRIVASTAVA, Indresh  
APPLICANT: LIAN, Yang  
APPLICANT: HARTOG, Karin  
APPLICANT: LIU, Hong  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475,515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 74  
LENGTH: 4689  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: SP160.modSP162.gag.modSP2  
US-09-475-515-74  
Query Match 81.0%; Score 1197.8; DB 4; Length 4689;  
Best Local Similarity 89.3%; Pred.No. 9.3e-177;  
Matches 1347; Conservative 0; Mismatches 132; Indels 30; Gaps 4;  
QY 1 ATGGGCGCGCGCGCAGCATCTGCGCGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
Db 3156 ATGGGCGCGCGCGCAGCATCTGCGCGCGGCAAGAGTACAGCTGAAGCATCTGTGGGCGCAGCGCGAG 3215  
QY 61 CTGCGCGCGCGCGCAAGTGTCTACATGATGAAGCATCTGTGTGGGCGCAGCGCGAG 120  
Db 3216 CTGCGCGCGCGCGCAAGAGTGTCTACATGATGAAGCATCTGTGTGGGCGCAGCGCGAG 3275  
QY 121 CTGAGAGATGTCGCTGTGAACCGCGCGCTGTGAGAGCAACGAGGCGTGAAGAGATC 180  
Db 3276 CTGAGAGATGTCGCTGTGAACCGCGCGCTGTGAGAGCAACGAGGCGTGAAGAGATC 3335  
QY 181 ATCCGCAAGCTGCAACCGCGCGCTGTGAGACCGGCAAGAGAGTGAAGAGCTGTTCAAC 240  
Db 3336 CTGCGCGCGCGCGCAAGCTGTGAACCGCGCGCTGTGAGAGCAACGAGGCGTGAAGAGATC 3395  
QY 241 ACCGTGCGCACCTGTGACTGCTGTGACAGAAATGAGTCCGCGACACCAAGAGGCGC 300  
Db 3396 ACCGTGCGCACCTGTGACTGCTGTGACAGAGCAATGCACTGAAGACCAAGAGAGGCGC 3455  
QY 301 CTGACAAAGATCGAGAGAGCAACAAGTCCAGAGCAAGATCCAGCGAGGCGC 360  
Db 3456 CTGAGAAAGATCGAGAGAGCAACAAGTCCAGAGCAAGATCCAGCGAGGCGCGCGC 3515  
QY 361 GCCG-----ACAAAGGCAAGGTGAGCAAGTCACTCCCATGTGCAAGACTG 408  
Db 3516 GCCGCGCGCACCGGCAACAGAGCAAGTGAAGCAAGTCACTCCCATGTGCAAGACTG 3575  
QY 409 CAGGCGCAGATGTGACACAGGCGCATCAAGCGCCCGCACTTGAAAGCTGTGAGAGTG 468  
Db 3576 CAGGCGCAGATGTGACACAGGCGCATCAAGCGCCCGCACTTGAAAGCTGTGAGAGTG 3635  
QY 469 ATCGAGAGAGAGCTTGAAGCGCGCGAGGTGATCCCATGTTCAACGCGCTGAGAGAGGCG 528  
Db 3636 GTGAGAGAGAGCTTGAAGCGCGCGAGGTGATCCCATGTTCAAGCGCTGAGAGAGGCG 3695  
QY 529 GCCACCCCCCAGAGACTGGAACAAGATGTTGAACACCGTGGGGGGGACCAAGGCGCGCATG 588

```
Db      3696  GCCACCCCCCAGAGACCTGGAACAAGATGTGAACACCGTGGGGGCGGACCAAGGCGCCGATG 3755
Qy      589  CAGATGCTGAAGAGACCATCAACGAGAGGCGCGGAGTGGGACCGCGTGAACCCCGG 648
Db      3756  CAGATGCTGAAGAGACCATCAACGAGAGGCGCGGAGTGGGACCGCGTGAACCCCGG 3815
Qy      649  CAGCGCGGCGCCATCGCCCCCGGCGAGATGCGGAGCCCCGGGCGAGCAATCGCGCG 708
Db      3816  CAGCGCGGCGCCATCGCCCCCGGCGAGATGCGGAGCCCCGGGCGAGCAATCGCGCG 3875
Qy      709  ACCACGAGACCTTGAGAGAGATGCGCTTGATGACGAGAACCCCTCCATCCCGTG 768
Db      3876  ACCACGAGACCTTGAGAGAGATGCGCTTGATGACGAGAACCCCTCCATCCCGTG 3935
Qy      769  GCGGACATTAAACGCGTGTGATCATCTCGGCGCTGAAACAATGTGTGGATGTAACG 828
Db      3936  GCGGACATTAAACGCGTGTGATCATCTCGGCGCTGAAACAATGTGTGGATGTAACG 3995
Qy      829  CCGGTGAGCATCTGGAACATCAAGAGGCGCCCAAGAGACCTTCGCGGACTACGTGAC 888
Db      3996  CCGGTGAGCATCTGGAACATCTGGAACATCTGGAACATCTGGAACATCTGGAAC 4055
Qy      889  CGCTTCTTCAAGACCTGCGCGCGCGAGAGAGACCCAGAGAGTGAAGAACTGATGAC 948
Db      4056  CGCTTCTTCAAGACCTGCGCGCGCGAGAGAGACCCAGAGAGTGAAGAACTGATGAC 4115
Qy      949  GACACCTGCTGTGTGAGAAAGCCCAACCCCGACTGCAACCAATCTGCGCGCTCTCGG 1008
Db      4116  GAGACCTGCTGTGTGAGAAAGCCCAACCCCGACTGCAACCAATCTGCGCGCTCTCGG 4175
Qy      1009  CCGGCGCGGCGCTGAGAGAGATGATGACCGGCTGCGAGGCGGCGGCGCCGAGCCAC 1068
Db      4176  CCGGCGCGGCGCTGAGAGAGATGATGACCGGCTGCGAGGCGGCGGCGGCGGCGGCG 4235
Qy      1069  AAGGCGCGGCTGCTGAGAGAGATGATGACCGGCGGCGGCGGCGGCGGCGGCGGCG 1122
Db      4236  AAGGCGCGGCTGCTGAGAGAGATGATGACCGGCGGCGGCGGCGGCGGCGGCGGCG 4295
Qy      1123  CAGAGAGAGCACTTCAAGAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1182
Db      4296  CAGCGCGGCACTTCCGCAACGAGCGGAGAACCGTCAAGTGTCTTCAACTGCGGCAAGG 4355
Qy      1183  GCGCAATGCGCGGCACTGCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1242
Db      4356  GCGCAATGCGCGGCACTGCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4415
Qy      1243  GAGGCGCACTGAGATGAGAGATGAGACCGAGCGGCGGCGGCGGCGGCGGCGGCGG 1302
Db      4416  GAGGCGCACTGAGATGAGAGATGAGACCGAGCGGCGGCGGCGGCGGCGGCGGCGG 4475
Qy      1303  CCGAGCGCACTGAGAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1362
Db      4476  CCGAGCGCACTGAGAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4535
Qy      1363  CCGCGCGGAGAGTTCGCGCTT-----GAGAGAGACCAACCCCGGCGGAGAGAG--- 1413
Db      4536  CCGCGCGGAGAGTTCGCGCTT-----GAGAGAGAGAGAGAGAGAGAGAGAGAG--- 4595
Qy      1414  ---AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1470
Db      4596  ATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4655
Qy      1471  AGCCAGTAA 1479
Db      4656  AGCCAGTAA 4664
```

RESULT 5  
US-09-475-515-73  
; Sequence 73, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:

```
APPLICANT: BARNETT, Susan
APPLICANT: ZUR MEGEDE, Jan
APPLICANT: SRIVASTAVA, Indresh
APPLICANT: LIAN, Ying
APPLICANT: HARTOG, Karin
APPLICANT: LIU, Hong
APPLICANT: GREER, Catherine
APPLICANT: SELBY, Mark
APPLICANT: WALKER, Christopher
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
FILE REFERENCE: 1621.002
CURRENT FILING DATE: 1999-12-30
NUMBER OF SEQ ID NOS: 90
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 73
LENGTH: 4766
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:
US-09-475-515-73
```

Query Match 81.0% Score 1197.8; DB 4; Length 4766;  
Best Local Similarity 89.3%; Pred. No. 9.3e-177;  
Matches 1347; Conservative 0; Mismatches 133; Indels 30; Gaps 4;

```
Qy      1  ATGGGCGCGCGCGGCGGAGATCTGCGCGCGGCGGCAAGCTGAGCGCTGGAGCGCATCCGC 60
Db      3233  ATGGGCGCGCGCGGCGGAGATCTGCGCGCGGCGGCGGCGGCGGCAAGCTGAGCGCG 3292
Qy      61  CTGCGCGCGCGCGGCGGAGAGTGTCTACATGATGAAGCACTGTGTGTGGCGCGCGGAG 120
Db      3293  CTGCGCGCGCGCGGCGGAGAGAGTGTCTACATGATGAAGCACTGTGTGTGGCGCGCGG 3352
Qy      121  CTGAGAGAGTTCGCGCTGACCGCGCGCTGCTGAGAGACGAGAGAGAGGCTGCAAGAGATC 180
Db      3333  CTGAGAGAGTTCGCGCTGACCGCGCGCTGCTGAGAGACGAGAGAGAGGCTGCGGAGATC 3412
Qy      181  ATCCGCGAGCTGCAACCGCGCGCTGCGAGACCGGCGAGAGAGAGTGTGAAGAGCTGTTCAAC 240
Db      3413  CTGCGCGAGCTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3472
Qy      241  ACCGTGCGCACCTGTACTGCGGTGCAAGAGATGAGTCCGCGACCAAGAGAGGCG 300
Db      3473  ACCGTGCGCACCTGTACTGCGGTGCAAGAGATGAGTCCGCGACCAAGAGAGGCGCG 3532
Qy      301  CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db      3533  CTGAGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3592
Qy      361  GCGG-----ACAAGGCGCAAGGTGAGCGCAAGTCAACCCCATGCTGTGAGAACTG 408
Db      3593  GCGGCGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3652
Qy      409  CAGGCGAGATGTGACACAGGCGATCGACCGCGCGGCGGCGGCGGCGGCGGCGGCGG 468
Db      3653  CAGGCGAGATGTGACACAGGCGATCGACCGCGCGGCGGCGGCGGCGGCGGCGGCGG 3712
Qy      469  ATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 528
Db      3713  GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3772
Qy      529  GCCACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
Db      3773  GCCACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3832
Qy      589  CAGATGCTGAAGAGACCATCAACGAGAGAGCGCGGCGGAGTGGAGCCGCGTGAACCCG 648
Db      3833  CAGATGCTGAAGAGAGACCATCAACGAGAGAGCGCGGCGGAGTGGAGCCGCGTGAAC 3892
```

QY 649 CACGCGGCGCCCATGCGCCCGGCGAGATGCGGAGAGCCCGCGGAGGACATGCGCGGC 708  
DB 3893 CACGCGGCGCCCATGCGCCCGGCGAGATGCGGAGAGCCCGCGGAGGACATGCGCGGC 3952  
QY 709 ACCACGAGACCCCTGAGAGAGAGATGCGCTGATGATGACGAGCAACCCCGCATCCCGCTG 768  
DB 3953 ACCACGAGACCCCTGAGAGAGAGATGCGCTGATGATGACGAGCAACCCCGCATCCCGCTG 4012  
QY 769 GCGGACATCTACAGAGCGGTGATCATCTGCGGCTGGAACAGATGTGCGATGTACAGC 828  
DB 4013 GCGGAGATCTACAGAGCGGTGATCATCTGCGGCTGGAACAGATGTGCGATGTACAGC 4072  
QY 829 CCGGAGAGATCTGAGATCATGAGAGAGAGCCCGGAGAGCCCTTCCGCGACTACGTGAGC 888  
DB 4073 CCGGAGAGATCTGAGATCATGAGAGAGAGCCCGGAGAGCCCTTCCGCGACTACGTGAGC 4132  
QY 889 CGCTTCTTCAAGACCTTGGCGCGCGAGAGAGAGCCCGGAGAGAGATGTAAGACTGATGAC 948  
DB 4133 CGCTTCTTCAAGACCTTGGCGCGCGAGAGAGAGCCCGGAGAGAGATGTAAGACTGATGAC 4192  
QY 949 GACACCTGCTGTGTGAGAGAGAGAGCCCGGAGAGAGATGTAAGACTGATGAGCCTTCCG 1008  
DB 4193 GAGAGCTGCTGTGTGAGAGAGAGAGCCCGGAGAGAGATGTAAGACTGATGAGCCTTCCG 4252  
QY 1009 CCGGCGCGAGCGTGAAGAGATGATGACCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 1068  
DB 4253 CCGGCGCGAGCGTGAAGAGATGATGACCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 4312  
QY 1069 AAGGCGCGGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1122  
DB 4313 AAGGCGCGGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4372  
QY 1123 CAGAGAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1182  
DB 4373 CAGAGAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4432  
QY 1183 GAGGCGCATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1242  
DB 4433 GAGGCGCATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4492  
QY 1243 GAGGCGCATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302  
DB 4493 GAGGCGCATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4552  
QY 1303 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1362  
DB 4553 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4612  
QY 1363 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1413  
DB 4613 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4672  
QY 1414 --AGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1470  
DB 4673 ATCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4732  
QY 1471 AGCCAGTAA 1479  
DB 4733 AGCCAGTAA 4741

## RESULT 6

US-09-475-515-7  
; Sequence 7, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine

APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475.515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 2031  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: HIV-Gag/HCV-core Fusion polypeptide  
US-09-475-515-7  
Query Match 80.9%; Score 1195.8; DB 4; Length 2031;  
Best Local Similarity 89.3%; Pred. No. 1.9e-176;  
Matches 1345; Conservative 0; Mismatches 132; Indels 30; Gaps 4;  
QY 1 ATGGGCGCGCGCGCGAGCATCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
DB 7 ATGGGCGCGCGCGCGAGCGTCTGAGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66  
QY 61 CTGGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 67 CTGGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126  
QY 121 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 127 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186  
QY 181 ATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 187 CTGGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
QY 241 ACCGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
DB 247 ACCGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
QY 301 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCGG-----ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 408  
DB 367 GCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426  
QY 409 CAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468  
DB 427 CAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486  
QY 469 ATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 528  
DB 487 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546  
QY 529 GCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588  
DB 547 GCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606  
QY 589 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648  
DB 607 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666  
QY 649 CACGCGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 708  
DB 667 CACGCGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
QY 709 ACCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 768  
DB 727 ACCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786



QY 769 GGGCATCTACAAAGGGTGGATCATCTGGGCTGAACAAGATCGTGGATGTACAGC 828  
| | | | |  
Db 787 GGGGATCTTACAAAGGGTGGATCATCTGGGCTGAACAAGATCGTGGATGTACAGC 846  
| | | | |  
QY 829 CCCGTGAGCATCTGAGCATCAAGAGGCCCCAAGAGCCCTCCGCACTACGTGAC 888  
| | | | |  
Db 847 CCCACAGCATCTGAGCATCTGAGGCCCCAAGAGCCCTCCGCACTACGTGAC 906  
| | | | |  
QY 889 CGCTTTTCAAGACCTTGGGCGCCGAGCAGAGCAACCAAGAGGTGAAGAACTGATGACC 948  
| | | | |  
Db 907 CGCTTTTCAAGACCTTGGGCGCCGAGCAGAGCAACCAAGAGGTGAAGAACTGATGACC 966  
| | | | |  
QY 949 GACACCTGCTGTGAGAACCGCAACCCCGACTGAAGCACTCTGCGCGCTCTGAC 1008  
| | | | |  
Db 967 GAGACCTGCTGTGAGAACCGCAACCCCGACTGAAGCACTCTGCGCGCTCTGAC 1026  
| | | | |  
QY 1009 CCCGCGCAGCATCTGAGAGATGATGACCGCTGCAAGGCGTGGGCGGCCCCAGCCAC 1068  
| | | | |  
Db 1027 CCCGCGCAGCATCTGAGAGATGATGACCGCTGCAAGGCGTGGGCGGCCCCAGCCAC 1086  
| | | | |  
QY 1069 AAGGCGCGGTGTGAGCGGAGGAGTACCGAGCCCAACA-----CAGCGTATGATG 1122  
| | | | |  
Db 1087 AAGGCGCGGTGTGAGCGGAGGAGTACCGAGCCCAACA-----CAGCGTATGATG 1146  
| | | | |  
QY 1123 CAGAGAGCACTTCAAGGCGCCCGCGCGCATGTCTCAAGTGTCTTCACTGCGGCAAGAG 1182  
| | | | |  
Db 1147 CAGCGCGGCACTTCCGCAACCAAGCGGAGACCGTCAAGTGTCTTCACTGCGGCAAGAG 1206  
| | | | |  
QY 1183 GGCACATGCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242  
| | | | |  
Db 1207 GGCACATGCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1266  
| | | | |  
QY 1243 GAGGCGCACAGATGAAGAGCTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1302  
| | | | |  
Db 1267 GAGGCGCACAGATGAAGAGCTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1326  
| | | | |  
QY 1303 CCCAGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1362  
| | | | |  
Db 1327 CCCAGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386  
| | | | |  
QY 1363 CCCGCGCAGAGCTTCCGCTT-----GAGAGAGCAACCCCGCGCGCGCGCGCGCGCG 1413  
| | | | |  
Db 1387 CCCGCGCAGAGCTTCCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1446  
| | | | |  
QY 1414 ---AGCAAGAGCGCGAGACCTGACCAAGCTGAGAGAGCTGCGCAAGCAAGCGCGCG 1470  
| | | | |  
Db 1447 ATCGAAGAGAGCTGACCCCGCGAGCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1506  
| | | | |  
QY 1471 AGCCAGT 1477  
| | | | |  
Db 1507 AGCCAGT 1513  
| | | | |

## RESULT 7

US-09-475-515-5  
; Sequence 5, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Yang  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30

NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1853  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-09-475-515-5  
Query Match 75.4%; Score 1114.6; DB 4; Length 1853;  
Best Local Similarity 85.8%; Pred. No. 6.6e-164;  
Matches 125; Conservative 0; Mismatches 184; Indels 30; Gaps 4;  
QY 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
| | | | |  
Db 7 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66  
| | | | |  
QY 61 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
| | | | |  
Db 67 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 126  
| | | | |  
QY 121 CTGAGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
| | | | |  
Db 127 CTGAGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186  
| | | | |  
QY 181 ATCCGCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
| | | | |  
Db 187 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246  
| | | | |  
QY 241 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
| | | | |  
Db 247 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306  
| | | | |  
QY 301 CTGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
| | | | |  
Db 307 CTGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
| | | | |  
QY 361 GCCG-----ACAGAGGCAAGGTGAGCGCAAGCTACCCCATGCTGAGAGAGCGT 408  
| | | | |  
Db 367 GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426  
| | | | |  
QY 409 CAGGCGCGAGTGTGACCAAGGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 468  
| | | | |  
Db 427 CAGGCGCGAGTGTGACCAAGGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486  
| | | | |  
QY 469 ATCGAGAGAGAGCGCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528  
| | | | |  
Db 487 GTGAGAGAGAGCGCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546  
| | | | |  
QY 529 GCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 588  
| | | | |  
Db 547 GCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606  
| | | | |  
QY 589 CAGATGCTGAAGAGACCATCAACGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648  
| | | | |  
Db 607 CAGATGCTGAAGAGACCATCAACGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666  
| | | | |  
QY 649 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708  
| | | | |  
Db 667 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726  
| | | | |  
QY 709 ACCACCAAGACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 768  
| | | | |  
Db 727 ACCACCAAGACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786  
| | | | |  
QY 769 GCGGAGATCTACAAAGCGGTGATCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828  
| | | | |  
Db 787 GCGGAGATCTACAAAGCGGTGATCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 846  
| | | | |  
QY 829 CCCGAGACATCTGAGAGATCAAGAGAGGCCCCAAGAGCCCTTCCGCGAGCTACCTGAGC 888  
| | | | |

Db 847 CCCACGACATCTTGAGATCCGCGACGAGGCCCCAAGAGCCCTTCCGACTAGCTGAC 906  
Qy 889 CGCTTTCTTCAAGACCTCTGCGCGCCGAGACACCCAGAGGTGAGAACTGATGAC 948  
Db 907 CGCTTCTAAGAACCTCTGCGCGCTGAGAGGCGACCGACGAGACGTGAAGAACTGATGAC 966  
Qy 949 GACACCTCTGCTGCTGAGAACGCGCAACCCCGACTGCAAGACCATCTCTGCGCTCTGCGC 1008  
Db 967 GAGACCTCTGCTGCTGAGAACGCGCAACCCCGACTGCAAGACCATCTCTGCGCTCTGCGC 1026  
Qy 1009 CCCGCGCGCAGCTGAGAGATGATGACCCGCTCCAGAGGCGTGGCGGCCCGACGAC 1068  
Db 1027 CCCGCGCGCAGCTGAGAGATGATGACCCGCTCCAGAGGCGTGGCGGCCCGACGAC 1086  
Qy 1069 AAGCGCGCTGCTGCTGAGAGGCGCGAGCGATGAGCCAGGCGCA-----CGAGCGTATGATG 1122  
Db 1087 AAGCGCGCTGCTGCTGAGAGGCGCGAGCGATGAGCCAGGCGCACTATGATG 1146  
Qy 1123 CAGAGAGCAACTTCAAGAGGCCCCCGCGCGCATCTCTCAAGTCTTCACTTGGCGAGAGAG 1182  
Db 1147 CAGCGCGCAACTTCCGCAACCGGAGAGACCTTCAAGTCTTCACTTGGCGAGAGAG 1206  
Qy 1183 GAGCGCATCTGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242  
Db 1207 GAGCGCATCTGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1266  
Qy 1243 GAGCGCGCACTGAGAGAGATGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1302  
Db 1267 GAGAGACACCAAAATGAAAGATGCACTGAGAGACAGGCGTAAATTTTGAAGAGATCTG 1326  
Qy 1303 CCCGCGCGCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1362  
Db 1327 CTTCTTCAAGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386  
Qy 1363 CCCGCGCGAGCTTCCGCTTCCAGAGAG-----ACCAACCGCGCGCGCGCGCGCGCGCG 1413  
Db 1387 CCAAGAGAGAGCTTCAAGGCTTGGGAGAGAGAAACAATCTCTCAAGAGAGAGAGCG 1446  
Qy 1414 ---AGCAAGAGAGCGGAGACCTGACCAAGCTGAAGAGCTGTTGCGACGAGCCCTG 1470  
Db 1447 ATAGACAGAGAACTATCTTTAATCTTCAATCTTCAATCATCTTGGGACAGACCCCTG 1506  
Qy 1471 AGCGAGTAA 1479  
Db 1507 TCACAGTAA 1515

RESULT 8  
US-09-475-515-78  
Sequence 78, Application US/09475515A  
Patent No. 6602705  
GENERAL INFORMATION:  
APPLICANT: BARNETT, Susan  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: SRIVASTAVA, Indresh  
APPLICANT: LIAN, Ying  
APPLICANT: HARTOG, Karin  
APPLICANT: LIU, Hong  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475,515A  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 78  
LENGTH: 1865  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: GPI  
US-09-475-515-78  
Query Match 75.4%; Score 1114.6; DB 4; Length 1865;  
Best Local Similarity 85.8%; Pred. No. 6.6e-164;  
Matches 1295; Conservative 0; Mismatches 184; Indels 30; Gaps 4;  
Qy 1 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 13 ATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72  
Qy 61 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 73 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 132  
Qy 121 CTGAGAGAGTTCCTGAGACCCCGCGCTGCTGAGACCGAGGCGCTGCAAGATC 180  
Db 133 CTGAGAGGCGCTTCGCGTGAACCCCGCGCTGCTGAGACCGAGGCGCTGCGCGCAAGTTC 192  
Qy 181 ATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 193 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 252  
Qy 241 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 253 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312  
Qy 301 CTGAGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 313 CTGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 372  
Qy 361 GCGC-----ACAGAGGCGAGGTGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 408  
Db 373 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432  
Qy 409 CAGGCGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468  
Db 433 CAGGCGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492  
Qy 469 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 528  
Db 493 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 552  
Qy 529 GCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 588  
Db 553 GCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612  
Qy 589 CAGATGCTGAAGAGACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648  
Db 613 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672  
Qy 649 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708  
Db 673 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 732  
Qy 709 ACCACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 768  
Db 733 ACCACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 792  
Qy 769 GCGACATCTCAAGAGCGGTGATCATCTGGGCTGGAACAAGATCTGCGAGTACAGC 828  
Db 793 GCGAGATCTCAAGAGCGGTGATCATCTGGGCTGGAACAAGATCTGCGAGTACAGC 852  
Qy 829 CCGGTGAGCATCTGAGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 888  
Db 853 CCGACAGCATCTGAGCATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 912  
Qy 889 CGCTTCTTCAAGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948  
Db 913 CGCTTCTTCAAGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 972  
Qy 949 GACACCTGCTGCTGAGAGAGCGCAACCCCGACTGCAAGACCATCTTGGCGCTCTGCGC 1008

Db 973 GAGACCTGTGTGTGAGAACGCAACCCCGACTGCAAGACCATCTCTAAGGCTCTCGGC 1032  
Qy 1009 CCCGCGCGCAGCTTGTGAGAGATGATGACCCGCTGCGAGGGGTGGGCGCCCGACCCAC 1068  
Db 1033 CCCGCGCGCAGCTTGTGAGAGATGATGACCCGCTGCGAGGGGTGGGCGCCCGACCCAC 1092  
Qy 1069 AAGGCGCGGTGTGTGCGAGGCGGATGAGACGAGGCCAACA-----CGAGCGTGAATGATG 1122  
Db 1093 AAGGCGCGGTGTGTGCGAGGCGGATGAGACGAGTGACCAACCCGCGAGCATCATGATG 1152  
Qy 1123 CAGAAAGCAACTTCAAGAGGCGCCCGCGCATGCTCAAGTGTCTTCAACTGCGGCAAGAG 1182  
Db 1153 CAGCGCGGAACTTCCGCAACCGAGGAAAGACCGTCAAGTGTCTTCAACTGCGGCAAGAG 1212  
Qy 1183 GCGCAATGCGCGGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242  
Db 1213 GCGCAATGCGCGGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1272  
Qy 1243 GAGGCGCGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302  
Db 1273 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1332  
Qy 1303 CCAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1362  
Db 1333 CTTCTCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1392  
Qy 1363 CCGCGCGAGAGCTTCCGCTTCAAGAGAG-----ACCAACCCCGCGCGAGAGAGAGAG--- 1413  
Db 1393 CAGAGAGAGAGCTTCAAGTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1452  
Qy 1414 ---AGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1470  
Db 1453 ATAGACAAAGAACTGTATCTTTAACTTCTCAAGTCACTTTGGCAAGAGAGAGAGAG 1512  
Qy 1471 AGCCAGTAA 1479  
Db 1513 TCACAGTAA 1521

RESULT 9  
US-09-475-515-79  
; Sequence 79, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Yang  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 79  
; LENGTH: 1865  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GP2  
US-09-475-515-79

Query Match 75.4%; Score 1114.6; DB 4; Length 1865;  
Best Local Similarity 85.8%; Pred. No. 6.6e-164;  
Matches 1295; Conservative 0; Mismatches 184; Indels 30; Gaps 4;

Qy 1 ATGGGCGCGCGCGCGAGCATCTGCGCGCGCGAGCAAGCTGAGACCCCTTGGAGAGCGCATCCGC 60

Db 13 ATGGGCGCGCGCGCGAGCGTGTGAGCGCGCGCGAGCTGAGCAAGTGGAGAGAAATCCGC 72  
Qy 61 CTGCGCGCGCGCGAGAGAGTCTTCAATGATGAAGCACTGTGTGTGGCCACCGCGAG 120  
Db 73 CTGCGCGCGCGCGAGAGAGTCAAGAGCTGAAGCAATGTGTGTGGCCACCGCGAG 132  
Qy 121 CTGAGAAATGTCCTGCTTAAACCCCGCGCTGTGAGACCAAGAGAGGCTGCAAGAGATC 180  
Db 133 CTGAGAGCGCTTGTGCGTGAACCCCGCGCTGTGAGACCAAGAGAGGCTGCGCAAGATC 192  
Qy 181 ATCCGCGAGTGCACCCCGCGCTGAGACCGCGAGCGAGAGAGTGAAGAGCTGTTCAC 240  
Db 193 CTGCGCGAGTGCACCCCGCGCTGAGACCGCGAGCGAGAGTGTGCGAGCTGTTCAC 252  
Qy 241 ACCGTGCGCACCTGTATCTGCTGTGACGAGAAAGATGAGGTTCGCGACCAAGAGAGCC 300  
Db 253 ACCGTGCGCACCTGTATCTGCTGTGACGAGCGCATGCTCAAGAGACCAAGAGAGCC 312  
Qy 301 CTGAGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 313 CTGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 372  
Qy 361 GCGG-----ACAGGGCAAGGTGAGCGCAGAACTACCCCATGTGCGAGAGAGCTG 408  
Db 373 GCGGCGCGCACCGGCAACAGAGCGAGCGGTGAGCGAGAGAGAGAGAGAGAGAGAGAGAG 432  
Qy 409 CAGGCGCAAGTGTGACACAGGCGCATAGCGCGCGCACCTGTGAAGCGCTGTGTAAGGTG 468  
Db 433 CAGGCGCAAGTGTGACACAGGCGCATAGCGCGCGCACCTGTGAAGCGCTGTGTAAGGTG 492  
Qy 469 ATCGAGAGAGAGAGCTTCAAGCGCGCGAGGTGATCCCATGTTCAACCGCGCTGAGAGAG 528  
Db 493 GTGAGAGAGAGAGAGCTTCAAGCGCGCGAGGTGATCCCATGTTCAAGCGCGCTGAGAGAG 552  
Qy 529 GCCACCGCGCGAGAGAGCTTGAACAGATGTTGAACAGGTGAGAGAGAGAGAGAGAGAGAG 588  
Db 553 GCCACCGCGCGAGAGAGCTTGAACAGATGTTGAACAGGTGAGAGAGAGAGAGAGAGAGAG 612  
Qy 589 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648  
Db 613 CAGATGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672  
Qy 649 CAGCGCGCGCGAGATGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 708  
Db 673 CAGCGCGCGCGAGATGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 732  
Qy 709 ACCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 768  
Db 733 ACCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 792  
Qy 769 GCGGAGATCTCAAGAGAGAGAGATCTGTGGGCTGAACAGAGAGAGAGAGAGAGAGAGAG 828  
Db 793 GCGGAGATCTCAAGAGAGAGAGATCTGTGGGCTGAACAGAGAGAGAGAGAGAGAGAGAG 852  
Qy 829 CCGGTGAGATCTTGAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 888  
Db 853 CCGGAGATCTTGAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 912  
Qy 889 GCGTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 948  
Db 913 GCGTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 972  
Qy 949 GACACCTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1008  
Db 973 GAGACCTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1032  
Qy 1009 CCGGCGCGCAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1068  
Db 1033 CCGGCGCGCAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1092  
Qy 1069 AAGGCGCGGTGTGTGCGAGGCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1122

Db 1093 AAGCCCGCGTGCTGGCCGAGGCGATGAGCAGATGAGCAACCCGCGACATCATGATG 1152  
Qy 1123 CAGAAAGCACTTCAAGGCCCCCGGCGCATGCTCAAGTCTTCAACTGCGGCAAGAG 1182  
Db 1153 CAGCCCGGCACTTCGCAACCGAGGAGACCGTCAAGTCTTCAACTGCGGCAAGAG 1212  
Qy 1183 GGCACATGCGCGCACTGCGCGCGCGCGCGCAAGAGGCGTGGCGAGTGGCGGCAAG 1242  
Db 1213 GGCACACCGCGCAAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1272  
Qy 1243 GAGGCGCAACAGATGAGGACTGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1302  
Db 1273 GAAAGCAACCAATGAAATGCTCACTGAGACAGGCTAATTTTATGAGGAATCTGG 1332  
Qy 1303 CCCAGCCACAAAGGCGCGCGCGCGCGCACTTCTGCAAGCGCGCGCGCGCGCGCG 1362  
Db 1333 CTTCTCTCAAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1392  
Qy 1363 CCGCGCGAGGCTTCGCGCTTCGAGGAG-----ACCACCGCGCGCGCGCGCGCG 1413  
Db 1393 CCGAAAGAGCTTCAAGGTTTGGGAGGAGAAACACTCTCTCAAGAGCGAGCGCG 1452  
Qy 1414 --AGCAAGAGCGCGAGACCTGACCAAGCTGAAAGCTGTCGCGCAAGCGCGCG 1470  
Db 1453 ATAGACAAAGGACTGATCTTTAACTTCACTCTCAAGTCACTCTTGGCAAGCGCG 1512  
Qy 1471 AGCCAGTAA 1479  
Db 1513 TCACAGTAA 1521

RESULT 10  
US-09-475-515-6  
; Sequence 6, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 4319  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: HIV-Gag-Polymerase  
US-09-475-515-6

Query Match 75.4%; Score 1114.6; DB 4; Length 4319;  
Best Local Similarity 85.8%; Pred No. 6.6e-164;  
Matches 1295; Conservative 0; Mismatches 184; Indels 30; Gaps 4;

Qy 1 ATGGGCGCGCGCGCGAGCATCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGCATCCGC 60  
Db 7 ATGGGCGCGCGCGCGAGCGTGTGAAGCGCGCGCGCGAGCTGAGCAAGTGGAGAAATCCGC 66  
Qy 61 CTGGCGCGCGCGCGCAAGAGTCTTACATGATGAAGCACTGGTGGCGCGCGCGAG 120  
Db 67 CTGGCGCGCGCGCGCAAGAGATCAAGCTGAAGCACTGTGGCGCGCGCGAG 126

Qy 121 CTGAGAGATTGCGCCCTGAAACCCCGCGCTGCTGAGAGACAGCGAGGCTGCAAGCAGATC 180  
Db 127 CTGAGCGCTTGGCCCTGTAACCCCGGCTTCTGAGAGACAGCGAGGCTGCGCGCAGATC 186  
Qy 181 ATCCGCGAGTTCACCCCGCGCTGCAAGCCGCGAGCGAGAGCTGMAAGGCTGTTCAAC 240  
Db 187 CTGGGCGAGCTGAGCGCGAGCGTGGCAAGCCGCGAGCGAGAGTGGCGAGGCTGTGAAC 246  
Qy 241 ACCGTGGCACCCCTGTATCTGCTGTGACAGAAAGTCCAGGTCGCGACACCAAGAGGCC 300  
Db 247 ACCGTGGCACCCCTGTATCTGCTGTGACAGAGGCTGACGTCGTCAGAACCAAGAGGCC 306  
Qy 301 CTGCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 307 CTGAGAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
Qy 361 GCGG-----ACAAAGGCAAGGTGAGCGCAAACTAACCCCATGTGCAAGACTG 408  
Db 367 GCGGCGGCGACCGGCGACAGAGCGAGGTGAGCGCAAGACTAACCCCATGTGCAAGACTG 426  
Qy 409 CAGGCGCGAGTGTGTCACACAGGCGCATGAGCCCGCGCACCTGAAAGCTTGGGTGAAGTG 468  
Db 427 CAGGCGCGAGTGTGTCACACAGGCGCATGAGCCCGCGCACCTGAAAGCTTGGGTGAAGTG 486  
Qy 469 ATCGAGAGAGAGGCTTCAAGCGCGAGGTGATCCCATGTTCAACGCGCTTGAAGCGAGGC 528  
Db 487 GTGAGAGAGAGGCTTCAAGCGCGAGGTGATCCCATGTTCAAGCGCTTGAAGCGAGGC 546  
Qy 529 GCCACCGCGCGAGACTTGAACAGATGTTGAACACCGTGGCGGCGCACAGGCGCGCATG 588  
Db 547 GCCACCGCGCGAGACTTGAACAGATGTTGAACACCGTGGCGGCGCACAGGCGCGCATG 606  
Qy 589 CAGATCTGAGAGAGACATCAAGAGAGAGGCGCGCGAGTGGAGCCGCGTGCACACCGCG 648  
Db 607 CAGATCTGAGAGAGAGACATCAAGAGAGAGGCGCGCGAGTGGAGCCGCGTGCACACCG 666  
Qy 649 CAGCGCGCGCGCGCGCGCGCGCGCGAGTGGAGGCGCGCGCGCGCGCGCGCGCGCG 708  
Db 667 CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726  
Qy 709 ACCACGAGACCTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 768  
Db 727 ACCACGAGACCTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786  
Qy 769 GCGGCACTTCAAGCGGTGATCATCTGGGCGCTGAAAGAGATCTGGAGATGATACGC 828  
Db 787 GCGGCACTTCAAGCGGTGATCATCTGGGCGCTGAAAGAGATCTGGAGATGATACGC 846  
Qy 829 CCGGTGAGCATCTGAGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 888  
Db 847 CCGGCGCGCATCTGAGCATCCGCGAGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 906  
Qy 889 CGTTTCTTCAAGACCTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 948  
Db 907 CGTTTCTTCAAGACCTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966  
Qy 949 GACACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1008  
Db 967 GAGACCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026  
Qy 1009 CCGGCGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1068  
Db 1027 CCGGCGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086  
Qy 1069 AAGGCGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1122  
Db 1087 AAGGCGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146  
Qy 1123 CAGAAAGCACTTCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1182  
Db 1147 CAGGCGCGCAACTTCCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206  
Qy 1183 GGCACATGCGCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1242

1207 GGGCCACCGCCAGGAACTGCGCCGCCCGCCCGCCAGAGAGGCTGCTGGCGCTGCGCGCCG 1266  
1243 GAGGGCCACGATGAGATGCTGCAACCGAGCGCCAGGCCAATCTCTGGGCAAGATCTGG 1302  
1267 GAAAGGACACCAATGAAAGATGCTGAGAGACAGGCTAATTTTATGGAGAAATCTGG 1326  
1303 CCCACGCAAGAGGCGCGCCGCAACTCTGCGAGAGCGCCGAGCGCCAGCGCCCGCC 1362  
1327 CTTCTTCAAGAGGAGGCGAGGAGGAAATTTTCTTCAAGAGAGACAGAGCCAGCCCGCA 1386  
1363 CCCGCGAGAGCTTCCGCTTTCAGAGG-----ACACCGCCGCGCGCAGAGAGAG--- 1413  
1387 CCAAGAGAGAGTTCAAGTTTGGGAGAGAGAAACAATCTCTCTCAAGAGAGAGCGCG 1446  
1414 --AGCAAGAGACCGGAGACCTGACCAAGCTGAAGAGCTGTTGGCAAGACCCCTG 1470  
1447 ATAGCAAGAACTATCTCTTAACTTCCCTCAGATCACTTTGGCAAGACCCCTCG 1506  
1471 AGCCGATA 1479  
1507 TCACAGTAA 1515

RESULT 11  
US-09-475-515-9  
Sequence 9, Application US/09475515A  
Patent No. 6602705  
GENERAL INFORMATION:  
APPLICANT: BARNETT, Susan  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: SRIVASTAVA, Indresh  
APPLICANT: LIAN, Ying  
APPLICANT: LIU, Hong  
APPLICANT: HARTOG, Karlin  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475,515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 9  
LENGTH: 1268  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag  
US-09-475-515-9

Query Match 69.3%; Score 1025.2; DB 4; Length 1268;  
Best Local Similarity 89.6%; Pred. No. 4, 2e-150;  
Matches 1131; Conservative 0; Mismatches 113; Indels 18; Gaps 2;

241 ACCGTGGCACTCTGTACTGGGTGACAGAGAAATCGAGGTCCGCGACCAAGAGGCGC 300  
247 ACCGTGGCACTCTGTACTGGGTGACAGAGCAATCGAGCTGAGAGCAACCAAGAGGCGC 306  
301 CTGACCAAGATCGAGAGAGACAGAACAAAGTCCAGCAGAAAGATTCAGAGGCGCAGGCGC 360  
307 CTGAGAGAAATCGAGAGAGACAGAACAAAGTCCAGAGAAAGAGGCGCAGAGGCGCGCGC 366  
361 GCGG-----ACAGAGGCAAGGTGAGCCAGAACTACCCCATGTGTGACAACTG 408  
367 GCGCGCGGCAACCGGCAACAGAGCGAGGTGAGCCAGAACTACCCCATGTGTGAGAACTG 426  
409 CAGGGCCAGAGTGTGACACAGAGCCATCAGCCCGGCAACCTGAAAGCGCTGGGTGAAGTG 468  
427 CAGGGCCAGAGTGTGACACAGAGCCATCAGCCCGGCAACCTGAAAGCGCTGGGTGAAGTG 486  
469 ATCGAGAGAAAGGCTTCAAGCCCGAGGTGATCCCATGTTCAACCGCCTGAGCAGAGGCGC 528  
487 GTGAGAGAGAAAGGCTTCAAGCCCGAGGTGATCCCATGTTCAAGCGCCTGAGCAGAGGCGC 546  
529 GCGACCCCGCAGAGCTTGAACAGATGTTGAACACCGTGGCGGCGCACAGGCGCGCATG 588  
547 GCGACCCCGCAGAGCTTGAACAGATGTTGAACACCGTGGCGGCGCACAGGCGCGCATG 606  
589 CAGATCTGAGAGACACCATGAACAGAGAGGCGCGGAGTGGAGACCGGCTGACCCCGTG 648  
607 CAGATCTGAGAGAGACCATGAACAGAGAGGCGCGGAGTGGAGACCGGCTGACCCCGTG 666  
649 CACGCGGCGCCCATTCGCGCGGCGAGATGCGCGAGCCCGCGGCGAGCAGACATCGCGGCG 708  
667 CACGCGGCGCCCATTCGCGCGGCGAGATGCGCGAGCCCGCGGCGAGCAGACATCGCGGCG 726  
709 ACCACAGACCTCTGACAGAGAGATGCGCTGATGACAGACACCCCGCATTCGCGGCG 768  
727 ACCACAGACCTCTGACAGAGAGATGCGCTGATGACAGACACCCCGCATTCGCGGCG 786  
769 GCGGCACTCTCAAGAGCGGTGATCATCTGGGCGCTGAAACAAGATGTGCGGATGTAACGC 828  
787 GCGGAGATCTCAAGAGCGGTGATCATCTGGGCGCTGAAACAAGATGTGCGGATGTAACGC 846  
829 CCGGTGACATCTGATCATGAGAGAGGCGCCCAAGAGCCCTTCCGAGATCACTGAGAC 888  
847 CCGACAGACATCTGATCATGAGAGAGGCGCCCAAGAGCCCTTCCGAGATCACTGAGAC 906  
889 CGCTTCTTCAAGACCTTGGCGGCGGCGAGACACCGAGAGGTGAAGAACTGATGACC 948  
907 CGCTTCTTCAAGACCTTGGCGGCGGCGAGACACCGAGAGGTGAAGAACTGATGACC 966  
949 GACACCTGCTGTGTGACAAAGCCCAACCCCGATGCAAGACCATCTGCGGCTCTCGGC 1008  
967 GAGACCTGCTGTGTGACAAAGCCCAACCCCGATGCAAGACCATCTGCGGCTCTCGGC 1026  
1009 CCGGCGCGCAGCTGTGAGAGATGATGACCGCTGCGAGGCGTGGGCGGCGCCAGCCAC 1068  
1027 CCGGCGCGCAGCTGTGAGAGATGATGACCGCTGCGAGGCGTGGGCGGCGGCGCCAC 1086  
1069 AAGGCGCGGCTGTGGCGGAGCGATGAGCCAGCA-----CAGCGTGTGATG 1122  
1087 AAGGCGCGGCTGTGGCGGAGCGATGAGCCAGGTGAGCAACCCGCGACCATCATGATG 1146  
1123 CAGAGAGCAACTTTCAGAGGCGCGCGGCGCATCTGCAAGTCTTCAACTTGGCGCAAGAG 1182  
1147 CAGCGCGGCAACTTTCAGAGGCGCGCGGCGCATCTGCAAGTCTTCAACTTGGCGCAAGAG 1206  
1183 GCGGCACTGCGCGGCAACTTGGCGGCGCGCGGCGCAAGAGGCTGTGAGAGTGGCGGCAAG 1242  
1207 GCGGCACTGCGCGGCAACTTGGCGGCGCGCGGCGCAAGAGGCTGTGAGAGTGGCGGCGC 1266  
1243 GA 1244  
1267 GA 1268















GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 12, 2005, 21:18:31 ; Search time 897.894 Seconds  
(without alignments)  
9798.345 Million cell updates/sec

Title: US-09-475-704A-3

Sequence: 1479  
1 atggggcggccgcgcagcat.....acgacccctgagccagtaaa 1479

Scoring table: IDENTITY NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 5537552 seqs, 2974263231 residues

Total number of hits satisfying chosen parameters: 11075104

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications NA:\*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
12:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
17:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
18:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
19:	/cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
20:	/cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1479	100.0	1479	10	US-09-967-464-63
2	1479	100.0	1479	10	US-09-899-575-3
3	1477.4	99.9	3162	15	US-10-190-435-18
4	1477.4	99.9	3462	15	US-10-190-435-16
5	1477.4	99.9	4419	15	US-10-190-435-19
6	1477.4	99.9	4419	16	US-10-190-305A-14
7	1477.4	99.9	4483	15	US-10-190-435-35
8	1477.4	99.9	4606	15	US-10-190-435-34
9	1477.4	99.9	4615	15	US-10-190-435-36
10	1477.4	99.9	4702	15	US-10-190-435-38
11	1477.4	99.9	4716	15	US-10-190-435-17

12	1477.4	99.9	4716	16	US-10-190-305A-13	Sequence 13, Appl
13	1475.8	99.8	2742	15	US-10-190-435-20	Sequence 20, Appl
14	1475.8	99.8	2742	16	US-10-190-305A-15	Sequence 15, Appl
15	1475.8	99.8	3930	15	US-10-190-435-9	Sequence 9, Appl
16	1475.8	99.8	3930	15	US-10-190-435-10	Sequence 10, Appl
17	1475.8	99.8	3930	15	US-10-190-435-11	Sequence 11, Appl
18	1475.8	99.8	5145	15	US-10-190-435-12	Sequence 12, Appl
19	1475.8	99.8	5145	16	US-10-190-305A-12	Sequence 12, Appl
20	1474.8	99.7	4713	15	US-10-190-435-59	Sequence 59, Appl
21	1474.8	99.7	4713	16	US-10-190-305A-83	Sequence 83, Appl
22	1473.8	99.6	5184	15	US-10-190-435-58	Sequence 58, Appl
23	1473.8	99.6	5184	16	US-10-190-305A-82	Sequence 82, Appl
24	1473.4	99.6	1479	10	US-09-967-464-67	Sequence 67, Appl
25	1472.4	99.6	2742	15	US-10-190-435-57	Sequence 57, Appl
26	1472.4	99.6	2742	16	US-10-190-305A-81	Sequence 81, Appl
27	1463	98.9	1479	10	US-09-899-575-20	Sequence 20, Appl
28	1315.2	88.9	3531	15	US-10-190-435-13	Sequence 13, Appl
29	1315.2	88.9	3537	15	US-10-190-435-14	Sequence 14, Appl
30	1315.2	88.9	3537	15	US-10-190-435-15	Sequence 15, Appl
31	1288.8	87.1	1491	10	US-09-899-575-51	Sequence 51, Appl
32	1288.8	87.1	1494	10	US-09-899-575-51	Sequence 51, Appl
33	1276.8	86.3	1509	10	US-09-899-575-4	Sequence 4, Appl
34	1275.2	86.2	1509	10	US-09-967-464-64	Sequence 64, Appl
35	1273.6	86.1	1509	10	US-09-967-464-68	Sequence 68, Appl
36	1260.8	85.2	1509	10	US-09-899-575-21	Sequence 21, Appl
37	1221.2	82.6	9166	17	US-10-359-120-168	Sequence 168, App
38	1206	81.5	9788	17	US-10-359-120-174	Sequence 174, App
39	1197.8	81.0	1515	17	US-10-387-336-4	Sequence 4, Appl
40	1197.8	81.0	4472	17	US-10-387-336-75	Sequence 75, Appl
41	1197.8	81.0	4608	17	US-10-387-336-76	Sequence 76, Appl
42	1197.8	81.0	4689	17	US-10-387-336-74	Sequence 74, Appl
43	1197.8	81.0	4766	17	US-10-387-336-73	Sequence 73, Appl
44	1196.2	80.9	2799	16	US-10-241-009-18	Sequence 18, Appl
45	1196.2	80.9	2799	16	US-10-190-434B-18	Sequence 18, Appl

## ALIGNMENTS

RESULT 1  
US-09-967-464-63  
Sequence 63, Application US/09967464  
Publication No. US20030138453A1  
GENERAL INFORMATION:  
APPLICANT: O'Hagan, Derek  
APPLICANT: O'Brien, Gillis  
APPLICANT: Donnelly, John J.  
APPLICANT: Polo, John M.  
APPLICANT: Barnett, Susan  
APPLICANT: Singh, Manohan  
APPLICANT: Ulmer, Jeffrey  
APPLICANT: Dubensky, Jr., Thomas W.  
TITLE OF INVENTION: MICROARTICLES FOR DELIVERY OF HETEROLOGOUS NUCLEIC ACIDS  
FILE REFERENCE: P16262.004  
CURRENT APPLICATION NUMBER: US/09/967,464  
CURRENT FILING DATE: 2002-04-11  
PRIOR APPLICATION NUMBER: 60/236,105  
PRIOR FILING DATE: 2000-09-28  
PRIOR APPLICATION NUMBER: 60/315,905  
PRIOR FILING DATE: 2001-08-30  
NUMBER OF SEQ ID NOS: 68  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 63  
LENGTH: 1479  
TYPE: DNA  
ORGANISM: Human immunodeficiency virus type 1  
US-09-967-464-63  
Query Match 100.0%; Score 1479; DB 10; Length 1479;  
Best Local Similarity 100.0%; Pred. No. 4.1e-308;  
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGGCGCCGCGCGCATCTCTGGCGGCGGCAAGCTGAGCGCTGGAGGCGCATCCGC 60

```
Db 1 ATGGGGGCGCCGCGCAGCATCTCTGGCGGGGCGCAAGCTGACCGCTCGGAGCGCATCCGC 60
Qy 61 CTGGCGCCCGCGCGCAAGAGTCTACATGATGAGACCTGTGTGGCCAGCCGCGAG 120
Db 61 CTGGCGCCCGCGCGCAAGAGTCTACATGATGAGACCTGTGTGGCCAGCCGCGAG 120
Qy 121 CTGGAGAAATTGGCCCTGAACCCCGGCTGTGAGAACCAAGAGGGGCTGCAAGCAATC 180
Db 121 CTGGAGAAATTGGCCCTGAACCCCGGCTGTGAGAACCAAGAGGGGCTGCAAGCAATC 180
Qy 181 ATCCGCACTGACACCCCGCTGTGACAGCCGCGAGAGAGCTGAAAGAGCTGTTCAAC 240
Db 181 ATCCGCACTGACACCCCGCTGTGACAGCCGCGAGAGAGCTGAAAGAGCTGTTCAAC 240
Qy 241 ACCGTGGCCACCTGTACTGCGTGCAGAGAAAGATTGAGGTCCGCGACCAAGAGGCC 300
Db 241 ACCGTGGCCACCTGTACTGCGTGCAGAGAAAGATTGAGGTCCGCGACCAAGAGGCC 300
Qy 301 CTGGACAAGATGAGAGGAGGAGCAAGAGTGCAGACAGATCCAGAGGCCGAGGCC 360
Db 301 CTGGACAAGATGAGAGGAGGAGCAAGAGTGCAGACAGATCCAGAGGCCGAGGCC 360
Qy 361 GCGCAAGAGGCAAGGTGAGCCAGAACTACCCCATCTGTCAGAACTTGAGGGCCAGATG 420
Db 361 GCGCAAGAGGCAAGGTGAGCCAGAACTACCCCATCTGTCAGAACTTGAGGGCCAGATG 420
Qy 421 GTGCAACAGGCAATCAGCCCGCGACCTGTAAAGCTGTGGTGAAGTATGAGAGAG 480
Db 421 GTGCAACAGGCAATCAGCCCGCGACCTGTAAAGCTGTGGTGAAGTATGAGAGAG 480
Qy 481 GCTTCAGCCCGGAGGTGATCCCATGTTCACCGGCTGAGGAGGGGCGCACCCCGAG 540
Db 481 GCTTCAGCCCGGAGGTGATCCCATGTTCACCGGCTGAGGAGGGGCGCACCCCGAG 540
Qy 541 GACCTGAACACGATGTTGAACAACCGTGGCGGCGCAACAGCCGCCATGACATGCTGAAG 600
Db 541 GACCTGAACACGATGTTGAACAACCGTGGCGGCGCAACAGCCGCCATGACATGCTGAAG 600
Qy 601 GACACCATTAAGAGAGGCGCGCGAGTGGAGACCGGTGTGACCCCGTGCAGCCCGGCC 660
Db 601 GACACCATTAAGAGAGGCGCGCGAGTGGAGACCGGTGTGACCCCGTGCAGCCCGGCC 660
Qy 661 ATCGCCCCCGGCAAGTGGCGAGCCCGCGGAGAGCATGCGCGGCAACCAAGCAC 720
Db 661 ATCGCCCCCGGCAAGTGGCGAGCCCGCGGAGAGCATGCGCGGCAACCAAGCAC 720
Qy 721 CTGCAAGAGCAGATGCGCTGATGACAGCAACCCCGCATCCCGTGGGCGACATCTAC 780
Db 721 CTGCAAGAGCAGATGCGCTGATGACAGCAACCCCGCATCCCGTGGGCGACATCTAC 780
Qy 781 AAGCGGTGATCATCTGTGGCTTGAACAAGATGTGCGGATGTACAGCCCGCTGAGCAATC 840
Db 781 AAGCGGTGATCATCTGTGGCTTGAACAAGATGTGCGGATGTACAGCCCGCTGAGCAATC 840
Qy 841 CTGGAATCAAGCAGAGGCGCCCAAGAGAGCCCTTCCGCGACTAGCTGAGCCGCTTTCTAC 900
Db 841 CTGGAATCAAGCAGAGGCGCCCAAGAGAGCCCTTCCGCGACTAGCTGAGCCGCTTTCTAC 900
Qy 901 ACCCTGGCGCGGAGAGAGCAACCAAGAGTGAAGAACTGATGACCGACACCTGTGTC 960
Db 901 ACCCTGGCGCGGAGAGAGCAACCAAGAGTGAAGAACTGATGACCGACACCTGTGTC 960
Qy 961 GTGCAAGAGCAGCAACCCCGACTGCAAGACATCTGCGGCTGTGGCCCGCGCGCAC 1020
Db 961 GTGCAAGAGCAGCAACCCCGACTGCAAGACATCTGCGGCTGTGGCCCGCGCGCAC 1020
Qy 1021 CTGGAAGAGATGATGACCGCTGCAAGAGGCGGTGGGCGCCCAAGCAAGGCGCGCGT 1080
Db 1021 CTGGAAGAGATGATGACCGCTGCAAGAGGCGGTGGGCGCCCAAGCAAGGCGCGCGT 1080
Qy 1081 CTGGCCGAGGCGATGAGCCAGGCGCAACAGCGTGTATGATGACAGAGAGCAATCTTCAAG 1140
Db 1081 CTGGCCGAGGCGATGAGCCAGGCGCAACAGCGTGTATGATGACAGAGAGCAATCTTCAAG 1140
```

```
Db 1081 CTGGCCGAGGCGATGAGCCAGGCGCAACAGCGTGTATGATGACAGAGCAATCTTCAAG 1140
Qy 1141 GGGCCCCCGCGCATGTGTCAAGTGTCTTCAACTGCGGCAAGAGGGGCGACATGCCCCGCAAC 1200
Db 1141 GGGCCCCCGCGCATGTGTCAAGTGTCTTCAACTGCGGCAAGAGGGGCGACATGCCCCGCAAC 1200
Qy 1201 TGCGCGCGCCCGCAAGAGGGCTGTGAAGTGGCGCAAGAGGGGCGACAGATGAAG 1260
Db 1201 TGCGCGCGCCCGCAAGAGGGCTGTGAAGTGGCGCAAGAGGGGCGACAGATGAAG 1260
Qy 1261 GACTGACACGAGCGCGAGGCAACTTCTGTGGCAAGATCTGAGCCAGCCAGCAAGGGCGCG 1320
Db 1261 GACTGACACGAGCGCGAGGCAACTTCTGTGGCAAGATCTGAGCCAGCCAGCAAGGGCGCG 1320
Qy 1321 CCGGCAACTTCTGTGCAAGCGCGCCGAGCCAGCCAGCCCGCGCGAAGCTTCCGC 1380
Db 1321 CCGGCAACTTCTGTGCAAGCGCGCCGAGCCAGCCAGCCCGCGCGAAGCTTCCGC 1380
Qy 1381 TTGAGAGAGCAACCCCGGCGCAAGAGGAGCAAGAGCCGCGAGACCTTGACACAGC 1440
Db 1381 TTGAGAGAGCAACCCCGGCGCAAGAGGAGCAAGAGCCGCGAGACCTTGACACAGC 1440
Qy 1441 CTGAAGAGCTGTTCGCAACAGACCCCTGAGCCAGTAA 1479
Db 1441 CTGAAGAGCTGTTCGCAACAGACCCCTGAGCCAGTAA 1479

RESULT 2
US-09-899-575-3
; Sequence 3, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
; APPLICANT: Zur Megele, Jan
; APPLICANT: Barnette, Susan W.
; APPLICANT: Egnelbrecht, Susan
; APPLICANT: van Rensburg, Estrelita Janse
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP01631.102
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 09/475,704
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic Gag
US-09-899-575-3

Query Match 100.0%; Score 1479; DB 10; Length 1479;
Best Local Similarity 100.0%; Pred. No. 4,1e-308;
Matches 1479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGGCGCGCGCGCAGCATCTCTGGCGGGGCGCAAGCTGAGCGCTGTGGAGCGCATCCGC 60
Db 1 ATGGGCGCGCGCGCAGCATCTCTGGCGGGGCGCAAGCTGAGCGCTGTGGAGCGCATCCGC 60
Qy 61 CTGGCGCCCGCGCGCAAGAGTCTACATGATGAGACCTGTGTGGCCAGCCGCGAG 120
Db 61 CTGGCGCCCGCGCGCAAGAGTCTACATGATGAGACCTGTGTGGCCAGCCGCGAG 120
Qy 121 CTGGAGAAATTGGCCCTGAACCCCGGCTGTGAGAACCAAGAGGGGCTGCAAGCAATC 180
Db 121 CTGGAGAAATTGGCCCTGAACCCCGGCTGTGAGAACCAAGAGGGGCTGCAAGCAATC 180
Qy 181 ATCCGCACTGACACCCCGCTGTGACAGCCGCGAGAGAGCTGAAAGAGCTGTTCAAC 240
Db 181 ATCCGCACTGACACCCCGCTGTGACAGCCGCGAGAGAGCTGAAAGAGCTGTTCAAC 240
```

QY 241 ACCGTGGCCCACTCTGTAATCTGCTGCAAGAGAAATGAGGTCCGCGACACCAAGAGGCC 300  
DB 241 ACCGTGGCCCACTCTGTAATCTGCTGCAAGAGAAATGAGGTCCGCGACACCAAGAGGCC 300  
QY 301 CTGGAACAATCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 301 CTGGAACAATCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 361 GCCGACAAAGGCGAAGGTGAGCCGAACTACCCCATCTGTGAGAACTTGCAAGGCCAGATG 420  
DB 361 GCCGACAAAGGCGAAGGTGAGCCGAACTACCCCATCTGTGAGAACTTGCAAGGCCAGATG 420  
QY 421 GTGCAACCAAGCCATCAGCCCCCGCCACCTTGAAAGCCTGGGTGAAGGTATCGAGAGAA 480  
DB 421 GTGCAACCAAGCCATCAGCCCCCGCCACCTTGAAAGCCTGGGTGAAGGTATCGAGAGAA 480  
QY 481 GCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTTGAGCGAGGGCGCACCCCCAG 540  
DB 481 GCTTCAGCCCCGAGGTGATCCCATGTTCAACGCGCTTGAGCGAGGGCGCACCCCCAG 540  
QY 541 GACTTGAACAGATGTTGAACACCGTGGCGGCGCACAGGCGCCATGCAAGATGCTGAAG 600  
DB 541 GACTTGAACAGATGTTGAACACCGTGGCGGCGCACAGGCGCCATGCAAGATGCTGAAG 600  
QY 601 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 601 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 661 ATGCGCCCCGCGAGATGCGCGAGCGCGCGAGGAGATGCGCGGAGCAACCAAGCACC 720  
DB 661 ATGCGCCCCGCGAGATGCGCGAGCGCGCGAGGAGATGCGCGGAGCAACCAAGCACC 720  
QY 721 CTGCAAGAGACAGATGCGCTGATGACAGCAACCCGCCATCCCGTGGCGCATCTAC 780  
DB 721 CTGCAAGAGACAGATGCGCTGATGACAGCAACCCGCCATCCCGTGGCGCATCTAC 780  
QY 781 AAGCGGTGATCATCTGGGCGCTGAAACAAGATGTGCGGATGTACAGCCCGTGAACATC 840  
DB 781 AAGCGGTGATCATCTGGGCGCTGAAACAAGATGTGCGGATGTACAGCCCGTGAACATC 840  
QY 841 CTGGAACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
DB 841 CTGGAACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 901 ACCCTGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 901 ACCCTGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
DB 961 GTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
QY 1021 CTGGAAGAGATGATGACCGCTGCGAGGCGGTGGGCGGCCCAAGAGAGAGAGAGAG 1080  
DB 1021 CTGGAAGAGATGATGACCGCTGCGAGGCGGTGGGCGGCCCAAGAGAGAGAGAGAG 1080  
QY 1081 CTGGAAGAGATGATGACCGCTGCGAGGCGGTGGGCGGCCCAAGAGAGAGAGAGAG 1140  
DB 1081 CTGGAAGAGATGATGACCGCTGCGAGGCGGTGGGCGGCCCAAGAGAGAGAGAGAG 1140  
QY 1141 GGGCGCGCGAGATGCTCAAGTGTCTCAACTGCGGCAAGAGAGAGAGAGAGAGAGAG 1200  
DB 1141 GGGCGCGCGAGATGCTCAAGTGTCTCAACTGCGGCAAGAGAGAGAGAGAGAGAGAG 1200  
QY 1201 TGGCGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
DB 1201 TGGCGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1261 GACTGCAACGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1261 GACTGCAACGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320

QY 1321 CCCGGCACTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
DB 1321 CCCGGCACTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
QY 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
QY 1441 CTGAAGAGCTGTTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
DB 1441 CTGAAGAGCTGTTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479

RESULT 3  
US-10-190-435-18  
; Sequence 18, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 3162  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagRtmuc\_C  
US-10-190-435-18

Query Match 99.9%; Score 1477.4; DB 15; Length 3162;  
Best Local Similarity 99.9%; Pred. No. 8-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCGCGCGCGAGAGATCTGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
DB 7 ATGGGCGCGCGCGAGAGATCTGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66  
QY 61 CTGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 67 CTGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126  
QY 121 CTGAAGAGATGAGCTTGAACCCCGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 180  
DB 127 CTGAAGAGATGAGCTTGAACCCCGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 186  
QY 181 ATCCGCGAGCTGACACCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 187 ATCCGCGAGCTGACACCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
QY 241 ACCGTGGCAACCTGTGATCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
DB 247 ACCGTGGCAACCTGTGATCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
QY 301 CTGGAACAATCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGGAACAATCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCCGACAAAGGCGAAGGTGAGCCGAACTACCCCATCTGTGAGAACTTGCAAGGCCAGATG 420  
DB 367 GCCGACAAAGGCGAAGGTGAGCCGAACTACCCCATCTGTGAGAACTTGCAAGGCCAGATG 426  
QY 421 GTGCAACAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
DB 427 GTGCAACAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486

481 GCCTTCAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTTGAAGGAGGCGCCACCCCCAG 540  
 487 GCTTTACAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTTGAAGGAGGCGCCACCCCCAG 546  
 541 GACCTGAACACGATGTTGAACACCGTGGCGGCGCACAGGCGCCATGAGATGCTGAAG 600  
 547 GACCTGAACACGATGTTGAACACCGTGGCGGCGCACAGGCGCCATGAGATGCTGAAG 606  
 601 GACACCATCAACGAGAGGCGCCGAGTGGAGCCGCTGACCCGCTGACCCGCGGCGCC 660  
 607 GACACCATCAACGAGAGGCGCCGAGTGGAGCCGCTGACCCGCTGACCCGCGGCGCC 666  
 661 ATGCGCCCCGCGCAGATGTCGACAGCCCGCGCAGAGATGCGCGGCGCACACGACACC 720  
 667 ATGCGCCCCGCGCAGATGTCGACAGCCCGCGCAGAGATGCGCGGCGCACACGACACC 726  
 721 CTGCAAGAGCAGATGCGCTGATGACCAACACCCGCCATCCGCTGGGCGACATCTAC 780  
 727 CTGCAAGAGCAGATGCGCTGATGACCAACACCCGCCATCCGCTGGGCGACATCTAC 786  
 781 AACCGGTGATCATCTGCGCTGAAACAAGATGTCGAGATGTAACGCCGCTGAGCATC 840  
 787 AACCGGTGATCATCTGCGCTGAAACAAGATGTCGAGATGTAACGCCGCTGAGCATC 846  
 841 CTGCAATCAACGAGGCGCCCAAGAGGCGCCCTTCCGGACTAGTGAACCGCTTTCTCAAG 900  
 847 CTGCAATCAACGAGGCGCCCAAGAGGCGCCCTTCCGGACTAGTGAACCGCTTTCTCAAG 906  
 901 ACCCTGCGCGCGCAGAGCAGACCAAGAGTGAAGACTGATGACCAACCTGCTG 960  
 907 ACCCTGCGCGCGCAGAGCAGACCAAGAGTGAAGACTGATGACCAACCTGCTG 966  
 961 GTGCAAGAGGCGCAACCCGCACTGCAAGACCATCTGCGCTTCTGCGCCCGCGCAC 1020  
 967 GTGCAAGAGGCGCAACCCGCACTGCAAGACCATCTGCGCTTCTGCGCCCGCGCAC 1026  
 1021 CTGGAAGAGATGATGACCGGCTGCAAGGCGCGTGGGCGCGCCAGCAAGGCGCGCG 1080  
 1027 CTGGAAGAGATGATGACCGGCTGCAAGGCGCGTGGGCGCGCCAGCAAGGCGCGCG 1086  
 1081 CTGCGCGAGCGATGAGCGGCGCAACACAGCGTATGATGACAGAGCAACTTCAAG 1140  
 1087 CTGCGCGAGCGATGAGCGGCGCAACACAGCGTATGATGACAGAGCAACTTCAAG 1146  
 1141 GGGCCCCGCGCATGCTCAAGTCTTCAACTGCGGCGAGAGGCGCATGCGCGCGCAC 1200  
 1147 GGGCCCCGCGCATGCTCAAGTCTTCAACTGCGGCGAGAGGCGCATGCGCGCGCAC 1206  
 1201 TGCGCGCGCGCGCAGAGAGGCGTGTGGAAGTGTGCGGCAAGAGGCGCGCAGATGAAG 1260  
 1207 TGCGCGCGCGCGCAGAGAGGCGTGTGGAAGTGTGCGGCAAGAGGCGCGCAGATGAAG 1266  
 1261 GACTGACCGAGCGCAGGCGCAACTTCTGCGCAAGATCTGCGCGACCAAGGCGCGC 1320  
 1267 GACTGACCGAGCGCAGGCGCAACTTCTGCGCAAGATCTGCGCGACCAAGGCGCGC 1326  
 1321 CCGCGCACTTCTCTCAAGGCGCGCGCAGCGCGCAACCGCGCGCGCGAGAGCTTCCGC 1380  
 1327 CCGCGCACTTCTCTCAAGGCGCGCGCAGCGCGCAACCGCGCGCGCGAGAGCTTCCGC 1386  
 1381 TTGCAAGAGACCAACCGCGCGCAGAGAGCAGAGCAAGAGCAAGAGCAAGAGCAAGAG 1440  
 1387 TTGCAAGAGACCAACCGCGCGCAGAGAGCAGAGCAAGAGCAAGAGCAAGAGCAAGAG 1446  
 1441 CTGAAGAGCTGTTGCGCAACGACCCCTGAGCCAGTAA 1479  
 1447 CTGAAGAGCTGTTGCGCAACGACCCCTGAGCCAGTAA 1485

RESULT 4  
 US-10-190-435-16  
 ; Sequence 16, Application US/10190435

Publication No. US20030143248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZUR MEGEDE, Jan  
 ; APPLICANT: BARNETT, Susan W.  
 ; APPLICANT: LIAN, Ying  
 ; APPLICANT: ENGELBRECHT, Susan  
 ; APPLICANT: VAN RENSBURG, Estrellita J.  
 ; TITLE OF INVENTION: POLYPEPTIDES ENCODING ANTIGENIC HIV TYPE C  
 ; FILE REFERENCE: PPI813.003 / 2302-18133  
 ; CURRENT APPLICATION NUMBER: US/10/190.435  
 ; CURRENT FILING DATE: 2002-12-30  
 ; NUMBER OF SEQ ID NOS: 319  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 16  
 ; LENGTH: 3462  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: GagProteinArtmut\_C  
 US-10-190-435-16  
 Query Match 99.9%; Score 1477.4; DB 15; Length 3462;  
 Best Local Similarity 99.9%; Pred. No. 8.8e-308;  
 Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 1 ATGGGCGCGCGCGCAGATCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGATCCGC 60  
 7 ATGGGCGCGCGCGCAGATCTGCGCGCGCGCAAGCTGAGCGCTGGAGCGATCCGC 66  
 61 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGATGGGCGCGCGCGAG 120  
 67 CTGCGCGCGCGCGCAAGAGTGTACATGATGAAGCACTGATGGGCGCGCGCGAG 126  
 121 CTGGAAGATGTCGCTGTAACCCGCGCTGCTGAGACCAAGAGGCTGCAAGCATC 180  
 127 CTGGAAGATGTCGCTGTAACCCGCGCTGCTGAGACCAAGAGGCTGCAAGCATC 186  
 181 ATCGCGCAGCTGCAACCCGCGCTGCAAGCCGCGCAGAGAGCTGGAAGGCTGTTCAAC 240  
 187 ATCGCGCAGCTGCAACCCGCGCTGCAAGCCGCGCAGAGAGCTGGAAGGCTGTTCAAC 246  
 241 ACCGTGCGCACCTCTGTACTGCGTGTGCAAGAGATTCAGAGTCCGCGACCAAGAGGCGC 300  
 247 ACCGTGCGCACCTCTGTACTGCGTGTGCAAGAGATTCAGAGTCCGCGACCAAGAGGCGC 306  
 301 CTGGAAGATGAGAGAGCAGAACTGTCAGCAGAGATTCAGAGGCGCGCGCGC 360  
 307 CTGGAAGATGAGAGAGCAGAACTGTCAGCAGAGATTCAGAGGCGCGCGCGC 366  
 361 GCCGCAAGAGGCGAGGTGAGCAACCTTCTGCGCAAGATTCAGAGGCGCGCGCGCAGATG 420  
 367 GCCGCAAGAGGCGAGGTGAGCAACCTTCTGCGCAAGATTCAGAGGCGCGCGCGCAGATG 426  
 421 GTGCAACGAGCGATCAGCGCGCGCAACCTTGAAGCTGAGAGTGAAGTGAAGAGAG 480  
 427 GTGCAACGAGCGATCAGCGCGCGCAACCTTGAAGCTGAGAGTGAAGTGAAGAGAG 486  
 481 GCTTTACAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTTGAAGGAGGCGCGCAC 540  
 487 GCTTTACAGCCCCGAGGTGATCCCATGTTTCAACCGCCCTTGAAGGAGGCGCGCAC 546  
 541 GACCTGAACACGATGTTGAACACCGTGGCGGCGCACAGGCGCCATGAGATGCTGAAG 600  
 547 GACCTGAACACGATGTTGAACACCGTGGCGGCGCACAGGCGCCATGAGATGCTGAAG 606  
 601 GACACCATCAACGAGAGGCGCCGAGTGGAGCCGCTGACCCGCTGACCCGCGGCGCC 660  
 607 GACACCATCAACGAGAGGCGCCGAGTGGAGCCGCTGACCCGCTGACCCGCGGCGCC 666  
 661 ATGCGCCCCGCGCAGATGTCGACAGCCCGCGCAGAGATGCGCGGCGCACACGACACC 720  
 667 ATGCGCCCCGCGCAGATGTCGACAGCCCGCGCAGAGATGCGCGGCGCACACGACACC 726

QY 721 CTGAGAGAGATGCGCTGGATGACCAACCCCACTCCCGTGGGCGACATCTAC 780  
DB 727 CTGAGAGAGAGATGCGCTGGATGACCAACCCCACTCCCGTGGGCGACATCTAC 786  
QY 781 AAGCGGTGATCATCTCTGGGCTTGAACAAAGATGTCGATGTACAGCCCGTGAACATC 840  
DB 787 AAGCGGTGATCATCTCTGGGCTTGAACAAAGATGTCGATGTACAGCCCGTGAACATC 846  
QY 841 CTGAGATCAAGAGAGGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 900  
DB 847 CTGAGATCAAGAGAGGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 906  
QY 901 ACCCTGGCGCGAGAGAGCAACCCAGAGGTGAAGAACTGATGACCAACCCGCTG 960  
DB 907 ACCCTGGCGCGAGAGAGCAACCCAGAGGTGAAGAACTGATGACCAACCCGCTG 966  
QY 961 GTGCAAGAACCCCACTGCAAGACCATCTCGCGCTCTCGGCGCGCGCGCAC 1020  
DB 967 GTGCAAGAACCCCACTGCAAGACCATCTCGCGCTCTCGGCGCGCGCGCAC 1026  
QY 1021 CTGAGAGAGATGATGACCCCTGTCAGAGGCGTGGCGCGCGCGCGCAC 1080  
DB 1027 CTGAGAGAGATGATGACCCCTGTCAGAGGCGTGGCGCGCGCGCGCAC 1086  
QY 1081 CTGGCGAGGCGATGAGCGAGCGCAACACAGGTGATGATGAGAGAGCACTTCAAG 1140  
DB 1087 CTGGCGAGGCGATGAGCGAGCGCAACACAGGTGATGATGAGAGAGCACTTCAAG 1146  
QY 1141 GCGCGCGCGCGATCTGCAAGTGTCTTCACTGCGCGAGAGAGGCGCACTCGCGCGCAC 1200  
DB 1147 GCGCGCGCGCGATCTGCAAGTGTCTTCACTGCGCGAGAGAGGCGCACTCGCGCGCAC 1206  
QY 1201 TCCCGCGCGCGCGAGAGAGGCTGCTGAGAGTGGCGCAAGAGGCGCACTGAGAG 1260  
DB 1207 TCCCGCGCGCGCGAGAGAGGCTGCTGAGAGTGGCGCAAGAGGCGCGCACTGAGAG 1266  
QY 1261 GACTGACCGAGCGCGAGCGCACTTCTGCGGCAAGATGTCGCGCGAGCGCAAGAGGCGCG 1320  
DB 1267 GACTGACCGAGCGCGAGCGCACTTCTGCGGCAAGATGTCGCGCGAGCGCAAGAGGCGCG 1326  
QY 1321 CCGCGCAACTTCTCTGAGAGCGCGCGCGAGCGCACTCGCGCGAGAGCTTCCG 1380  
DB 1327 CCGCGCAACTTCTCTGAGAGCGCGCGCGAGCGCACTCGCGCGAGAGCTTCCG 1386  
QY 1381 TTGAGAGAGCAACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1387 TTGAGAGAGCAACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
QY 1441 CTGAAGAGCTTGTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
DB 1447 CTGAAGAGCTTGTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1485

RESULT 5  
US-10-190-435-19  
; Sequence 19, Application US/10190435  
; General Information:  
; Application No. US20030143248A1  
; Applicant: ZUR MEGEDE, Jan  
; Applicant: BARNETT, Susan W.  
; Applicant: LIAN, Ying  
; Applicant: ENGELBRECHT, Susan  
; Applicant: VAN RENSBURG, Estrelita J.  
; Title of Invention: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; Title of Invention: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; File Reference: P/18133.003 / 2302-18133  
; Current Application Number: US/10/190,435  
; Number of Seq ID NOS: 319  
; Software: PatentIn Ver. 2.0  
; Seq ID No 19  
; Length: 4419

TYPE: DNA  
ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagPmtTatRevNef\_C  
US-10-190-435-19  
Query Match 99.9%; Score 1477.4; DB 15; Length 4419;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGGGCGCGCGCGAGAGATCTTGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
DB 7 ATGGGCGCGCGCGAGAGATCTTGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66  
QY 61 CTGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 67 CTGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126  
QY 121 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 127 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186  
QY 181 ATCCGCGAGCTGCAACCCCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 187 ATCCGCGAGCTGCAACCCCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
QY 241 ACCGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
DB 247 ACCGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
QY 301 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 307 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
QY 361 GCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
DB 367 GCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426  
QY 421 GTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
DB 427 GTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486  
QY 481 GCTTCAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
DB 487 GCTTCAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546  
QY 541 GACTGAAACAGATGTTGAACACCGTGGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
DB 547 GACTGAAACAGATGTTGAACACCGTGGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606  
QY 601 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 607 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666  
QY 661 ATCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
DB 667 ATCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
QY 721 CTGAGAGAGAGATGCGCTGGATGACCAACCCCACTCCCGTGGGCGACATCTAC 780  
DB 727 CTGAGAGAGAGATGCGCTGGATGACCAACCCCACTCCCGTGGGCGACATCTAC 786  
QY 781 AAGCGGTGATCATCTCTGGGCTTGAACAAAGATGTCGATGTACAGCCCGTGAACATC 840  
DB 787 AAGCGGTGATCATCTCTGGGCTTGAACAAAGATGTCGATGTACAGCCCGTGAACATC 846  
QY 841 CTGAGATCAAGAGAGGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 900  
DB 847 CTGAGATCAAGAGAGGCGCCCAAGAGCCCTTCGCGACTACGTGACCGCTTCTTCAAG 906  
QY 901 ACCCTGGCGCGAGAGAGCAACCCAGAGGTGAAGAACTGATGACCAACCCGCTG 960  
DB 907 ACCCTGGCGCGAGAGAGCAACCCAGAGGTGAAGAACTGATGACCAACCCGCTG 966





QY	1201	TGCGCGCCCCCGGAAGAGGGCTGTGGAAAGTCGCGCAAGAGAGGCACCAATGAAG	1260
Db	1207	TGCGCGCCCCCGGAAGAGGGCTGTGGAAAGTCGCGCAAGAGAGGCACCAATGAAG	1266
QY	1261	GACTGCACCGAGCGCCAGGCCCACTTCTGTGGCAAGATCTGGCCACGCCACAGAGGCCG	1320
Db	1267	GACTGCACCGAGCGCCAGGCCCACTTCTGTGGCAAGATCTGGCCACGCCACAGAGGCCG	1326
QY	1321	CCCGGCAACTTCTGTGCAAGAGCGCGCGCGAGGCCACCGCGCCCCCGCGCGAAGCTTCCG	1380
Db	1327	CCCGGCAACTTCTGTGCAAGAGCGCGCGCGAGGCCACCGCGCCCCCGCGCGAAGCTTCCG	1386
QY	1381	TTTCGAGAGACCAACCCCGCGCCAGAGCAAGAGAGCAAGACCGCGAAGCTTGAACGACG	1440
Db	1387	TTTCGAGAGACCAACCCCGCGCCAGAGCAAGAGAGCAAGACCGCGAAGCTTGAACGACG	1446
QY	1441	CTGAAAGAGCTTGTGGGCAAGACCCCTTGAAGCCGATGA	1479
Db	1447	CTGAAAGAGCTTGTGGGCAAGACCCCTTGAAGCCGATGA	1485

## RESULT 7

```

US-10-190-435-35
Sequence 35, Application US/10190435
Publication No. US20030143248A1
GENERAL INFORMATION:
APPLICANT: ZUR MEGEDE, Jan
APPLICANT: BARNETT, Susan W.
APPLICANT: LIAN, Ying
APPLICANT: ENGELBRECHT, Susan
APPLICANT: VAN RENSBURG, Estrellita J.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE
FILE REFERENCE: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
CURRENT APPLICATION NUMBER: US/10/190,435
CURRENT FILING DATE: 2002-12-30
NUMBER OF SEQ ID NOS: 319
SOFTWARE: PatentIn Ver. 2.0
Seq ID NO 35
LENGTH: 4483
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:
OTHER INFORMATION: gp160mod.tvl.dv1v2-gagmod.bw95
US-10-190-435-35

```

Query Match	Score	DB	Length
99.9%	1477.4	15	4483

Best Local Similarity 99.9%; Pred. No. 8.7e-308;

Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	ATGGGCGCCGCGCAGCATCTCTGGCGCGCGCAAGCTGGAGCGCTGGGAGGCGATCCGC	60
Db	3005	ATGGGCGCCGCGCAGCATCTCTGGCGCGCGCAAGCTGGAGCGCTGGGAGGCGATCCGC <td>306</td>	306
Qy	61	CTGGCGCCCGCGCGCAAGAGTCTACATGATGAAGCACCTTGSTGTGGCCACGCCGCGAG	120
Db	3065	CTGGCGCCCGCGCGCAAGAGTCTACATGATGAAGCACCTTGSTGTGGCCACGCCGCGAG	3122
Qy	121	CTGGGAAGTTGCGCCTTGAACTCCCGGCTGCTGGAGACCAAGCGAGGCTGGCAAGCATC	180
Db	3125	CTGGGAAGTTGCGCCTTGAACTCCCGGCTGCTGGAGACCAAGCGAGGCTGGCAAGCATC	318
Qy	181	ATCGCGCAGCTGCACCCCGCCCTGAGAACCGGCACGAGGAACTGAAGAGCTGTTTCAAC	240
Db	3185	ATCGCGCAGCTGCACCCCGCCCTGAGAACCGGCACGAGGAACTGAAGAGCTGTTTCAAC	3244
Qy	241	ACCGTGGCCACCTCTGTACTGCTGTCAGGAAATCGAGTCCGCGACACCAAGAGGCGC	300
Db	3245	ACCGTGGCCACCTCTGTACTGCTGTCAGGAAATCGAGTCCGCGACACCAAGAGGCGC	3300
Qy	301	CTGCACAAAGATCGAGAGGAGCGAACAAGTCCACGACAGAGATCCAGCAGGCCCGAGGCC	360

Db	3305	CTGGACAAAGATCGAGGAGGAGCAACAAGTGCAGACGAAGATCTCAGACGCGAGGCC	3364
Qy	361	GCCGACAAAGGCGAAGGTGAGCCAGAACTACCTCATCTGTGCAGAACCTTCAGAGGCGCAATG	420
Db	3365	GCCGACAAAGGCGAAGGTGAGCCAGAACTACCTCATCTGTGCAGAACCTTCAGAGGCGCAATG	3422
Qy	421	GTGACCAAGGCGATCAGACCCCGGACACCTTGAAAGCGCTGGGGTGAAGGTGATGAGAGAAAG	480
Db	3425	GTGACCAAGGCGATCAGACCCCGGACACCTTGAAAGCGCTGGGGTGAAGGTGATGAGAGAAAG	3484
Qy	481	GCTTTCAGACCCCGGAGGTATCCCATGTTCAACCCGCTGAGCGAGGCGCAACCCCGAG	540
Db	3485	GCTTTCAGACCCCGGAGGTATCCCATGTTCAACCCGCTGAGCGAGGCGCAACCCCGAG	3544
Qy	541	GACCTGAACAGATGTTGAACACGCTGGGCGGCGACCAAGCGCGCATGACAGATCTGAAG	600
Db	3545	GACCTGAACAGATGTTGAACACGCTGGGCGGCGACCAAGCGCGCATGACAGATCTGAAG	3604
Qy	601	GACACCATCAACGAGGAGGCGCGGATGGGACCGCGGTGCACCCCGTGCACGCGGCGCC	660
Db	3605	GACACCATCAACGAGGAGGCGCGGATGGGACCGCGGTGCACCCCGTGCACGCGGCGCC	3664
Qy	661	ATGCGCCCGGCGCAGATGCGCGAGCCCGCGGCGCGACATGCGCGGCGACCAAGCAGCAC	720
Db	3665	ATGCGCCCGGCGCAGATGCGCGAGCCCGCGGCGCGACATGCGCGGCGACCAAGCAGCAC	3722
Qy	721	CTGACGAGACGATGCGCTGTGATGACCAACACCCCGCATCCCGTGGGCGAATCTAC	780
Db	3725	CTGACGAGACGATGCGCTGTGATGACCAACACCCCGCATCCCGTGGGCGAATCTAC	3784
Qy	781	AAGCGGTGGATCATCTCTGGGCTTGAAACAAGATCGTGGGATGTACAGCCCGTGAACATC	840
Db	3785	AAGCGGTGGATCATCTCTGGGCTTGAAACAAGATCGTGGGATGTACAGCCCGTGAACATC	3844
Qy	841	CTGGACATCAAGCAGGCGGCCCAAGAGAACCTTCCGCACTACGAGAACCGCTTCTTCAAG	900
Db	3845	CTGGACATCAAGCAGGCGGCCCAAGAGAACCTTCCGCACTACGAGAACCGCTTCTTCAAG	3904
Qy	901	ACCCTGGCGCCCGAGCAGAGACCAACGAGGTGAAGAATTGGATGACCGACACCTGCTG	960
Db	3905	ACCCTGGCGCCCGAGCAGAGACCAACGAGGTGAAGAATTGGATGACCGACACCTGCTG	3964
Qy	961	GTGCGAAGCGCAACCCCGACTGCAAGCAATCTGCGCGCTCTTGGCCCGCGCGCAC	1020
Db	3965	GTGCGAAGCGCAACCCCGACTGCAAGCAATCTGCGCGCTCTTGGCCCGCGCGCAC	4024
Qy	1021	CTGAGAGGAGTGAATGACCGGCTGCGCAGGGGTGGGCGGCCCGACGCAAGAGCCCGGTG	1080
Db	4025	CTGAGAGGAGTGAATGACCGGCTGCGCAGGGGTGGGCGGCCCGACGCAAGAGCCCGGTG	4084
Qy	1081	CTGGCCGAGCGGATGAGCCAGGCGCAACACCGACGATGATGTGCAAGAGCGCAATTCAAG	1140
Db	4085	CTGGCCGAGCGGATGAGCCAGGCGCAACACCGACGATGATGTGCAAGAGCGCAATTCAAG	4144
Qy	1141	GCGCCCGCGCGCATGTCAGATGCTTCAACTGCGGCAAGAGGCGCAATGCGCGCAAC	1200
Db	4145	GCGCCCGCGCGCATGTCAGATGCTTCAACTGCGGCAAGAGGCGCAATGCGCGCAAC	4204
Qy	1201	TGCGCGCGCCCCCGGCAAGAGGGGTGTGGAAGTGCGGCAAGAGGGCGCAACATGAAG	1260
Db	4205	TGCGCGCGCCCCCGGCAAGAGGGGTGTGGAAGTGCGGCAAGAGGGCGCAACATGAAG	4264
Qy	1261	GACTGCAACGAGCGCCAGGCGCAACTTCTGTGGGCAAGATCTGCGCCACCAAGAGGCGCG	1320
Db	4265	GACTGCAACGAGCGCCAGGCGCAACTTCTGTGGGCAAGATCTGCGCCACCAAGAGGCGCG	4324
Qy	1321	CCCGGCAACTTCTGTGAGAGCGCGCCCGGACCGACCGCGCCCGCGCGAGGCTTCGCG	1380
Db	4325	CCCGGCAACTTCTGTGAGAGCGCGCGCCCGGACCGACCGCGCCCGCGCGAGGCTTCGCG	4384
Qy	1381	TTTGAAGAGACCAACCCCGGCGCAAGACAGAGACAGGACCGCGAGACCTTGACCAAG	1440
Db	4385	TTTGAAGAGACCAACCCCGGCGCAAGACAGAGACAGGACCGCGAGACCTTGACCAAG	4444

QY 1441 CTGAAGAGCTGTTCGGCAAGACCCCTGAGCCAGTAA 1479  
DB 4445 CTGAAGAGCTGTTCGGCAAGACCCCTGAGCCATATA 4483

RESULT 8  
US-10-190-435-34  
; Sequence 34, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARRETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 34  
; LENGTH: 4606  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: gp160mod.TV1.dvl.gagmod.BM965  
US-10-190-435-34

Query Match 99.9%; Score 1477.4; DB 15; Length 4606;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCGCCGCGGCGCAAGATCTGTGGCGGCGCAAGCTGAGCGCTGGAGGCGATCCGC 60  
DB 3128 ATGGGCGCCGCGGCGCAAGATCTGTGGCGGCGCAAGCTGAGCGCTGGAGGCGATCCGC 3187

QY 61 CTGGCGCCGCGGCGCAAGAGTCTACATGATGAAAGCACTGTGTGTGGCGCGCGAG 120  
DB 3188 CTGGCGCCGCGGCGCAAGAGTCTACATGATGAAAGCACTGTGTGTGGCGCGCGAG 3247

QY 121 CTGGAAGATTGCTGCTGAACCCCGGCTGTGTGAGACCAAGGAGCTGCAAGCATC 180  
DB 3248 CTGGAAGATTGCTGCTGAACCCCGGCTGTGTGAGACCAAGGAGCTGCAAGCATC 3307

QY 181 ATCCGCGAGCTGACCCCGGCTGTGAGACCGGAGGAGGAGCTGAAGAGCTGTTCAC 240  
DB 3308 ATCCGCGAGCTGACCCCGGCTGTGAGACCGGAGGAGGAGCTGAAGAGCTGTTCAC 3367

QY 241 ACCGTGGCCACCTGTACTGTGCTGACGAGAAAGATCGAGTCGCGACCAAGAGAGCC 300  
DB 3368 ACCGTGGCCACCTGTACTGTGCTGACGAGAAAGATCGAGTCGCGACCAAGAGAGCC 3427

QY 301 CTGGAAGATTGAGAGAGAGCAAGAAAGTGCACAGAAATCCAGACGCGCGAGCC 360  
DB 3428 CTGGAAGATTGAGAGAGAGCAAGAAAGTGCACAGAAATCCAGACGCGCGAGCC 3487

QY 361 GCGGCAAGAGGAGAGTGAAGCAAGATCCCATGTGCAAGAACTGAGAGGCGAGATG 420  
DB 3488 GCGGCAAGAGGAGAGTGAAGCAAGATCCCATGTGCAAGAACTGAGAGGCGAGATG 3547

QY 421 GTGCAACGAGCATCAAGCCCGGCAACCTGAACGCTGTGATGAAGTGAATCGAGAGA 480  
DB 3548 GTGCAACGAGCATCAAGCCCGGCAACCTGAACGCTGTGATGAAGTGAATCGAGAGA 3607

QY 481 GCGTTTACGCGCGAGAGTATTCCTCATGTTTCAACGCGCTTGAAGGAGGCGCACCCCG 540  
DB 3608 GCGTTTACGCGCGAGAGTATTCCTCATGTTTCAACGCGCTTGAAGGAGGCGCACCCCG 3667

QY 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCGCACAGGCGCATGAGATGCTGAAG 600

DB 3668 GACCTGAACACGATGTTGAACACCGTGGGCGGCGCACAGGCGCATGAGATGCTGAAG 3727  
QY 601 GACACCATCAACGAGAGAGCGCGAGTGGAGACCGGTGACCCCTGTGACGCGGCGCC 660  
DB 3728 GACACCATCAACGAGAGAGCGCGAGTGGAGACCGGTGACCCCTGTGACGCGGCGCC 3787

QY 661 ATCGCCCCCGGCGAGATGCGGAGCCCGCGGCAAGCATGCGCGGAGCAACGACGACC 720  
DB 3788 ATCGCCCCCGGCGAGATGCGGAGCCCGCGGCAAGCATGCGCGGAGCAACGACGACC 3847

QY 721 CTGCAAGAGAGATGCTGTGATGACAGCAACCCCATCTCCGTGGGAGCATCTAC 780  
DB 3848 CTGCAAGAGAGATGCTGTGATGACAGCAACCCCATCTCCGTGGGAGCATCTAC 3907

QY 781 AAGCGGTGATCATCTGTGGGCTTGAACAAAGTCTGCGGATGTAACGCCGTGAGCATC 840  
DB 3908 AAGCGGTGATCATCTGTGGGCTTGAACAAAGTCTGCGGATGTAACGCCGTGAGCATC 3967

QY 841 CTGACATCAAGCAAGGCGCCCAAGAGCCCTTCCGCGATACGTTGAGACCGCTTTCAG 900  
DB 3968 CTGACATCAAGCAAGGCGCCCAAGAGCCCTTCCGCGATACGTTGAGACCGCTTTCAG 4027

QY 901 ACCCTGCGCGCGAGCAAGACCAAGAGGTGAAGAACTGATGACCAACCTGTGCTG 960  
DB 4028 ACCCTGCGCGCGAGCAAGACCAAGAGGTGAAGAACTGATGACCAACCTGTGCTG 4087

QY 961 GTGCAAGACGCAACCCGCACTGCAAGCAATCTGTGCGGCTGTGCGCGCGCGGCGCAC 1020  
DB 4088 GTGCAAGACGCAACCCGCACTGCAAGCAATCTGTGCGGCTGTGCGCGCGCGGCGCAC 4147

QY 1021 CTGAGAGAGATGATGACCGGCTGCAAGGCGGTGGGCGGCGCGCACCAAGGCGCGG 1080  
DB 4148 CTGAGAGAGATGATGACCGGCTGCAAGGCGGTGGGCGGCGCGCACCAAGGCGCGG 4207

QY 1081 CTGGCGCGAGATGAGCGCAAGCAACAGCGTGTATGATGAGAAAGCAACTTCAAG 1140  
DB 4208 CTGGCGCGAGATGAGCGCAAGCAACAGCGTGTATGATGAGAAAGCAACTTCAAG 4267

QY 1141 GCGCCCGGCGCATCTGTCAAGTCTTCAACTGCGGCAAGAGGCGCACATGCGCGCAAC 1200  
DB 4268 GCGCCCGGCGCATCTGTCAAGTCTTCAACTGCGGCAAGAGGCGCACATGCGCGCAAC 4327

QY 1201 TGCGCGCGCGCGCAAGAGGCGTGTGAAGTGTGGGCGGCGCAAGAGGCGCAACGATGAG 1260  
DB 4328 TGCGCGCGCGCGCAAGAGGCGTGTGAAGTGTGGGCGGCGCAAGAGGCGCAACGATGAG 4387

QY 1261 GACTGCAACGAGCGCGCAAGCACTTCTGTGGCAAGATCTGCGCGCAACAGCAAGGCGCG 1320  
DB 4388 GACTGCAACGAGCGCGCAAGCACTTCTGTGGCAAGATCTGCGCGCAACAGCAAGGCGCG 4447

QY 1321 CCGGCAACTTCTGTGAGAGCGCGCGGAGCCGACCGCGCGCGCGGAGAGCTTCCG 1380  
DB 4448 CCGGCAACTTCTGTGAGAGCGCGCGGAGCCGACCGCGCGCGCGGAGAGCTTCCG 4507

QY 1381 TTGCAAGAGACCAACCCCGGCGCAAGAGAGGAGCAAGACCGGAGACCTTGAACGAGC 1440  
DB 4508 TTGCAAGAGACCAACCCCGGCGCAAGAGAGGAGCAAGACCGGAGACCTTGAACGAGC 4567

QY 1441 CTGAAGAGCTGTTCGGCAAGACCCCTGAGCCAGTAA 1479  
DB 4568 CTGAAGAGCTGTTCGGCAAGACCCCTGAGCCATATA 4606

RESULT 9  
US-10-190-435-36  
; Sequence 36, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARRETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan

```

; APPLICANT: VAN RENSBURG, Batrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 4615
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: sp16mod.TV1.dv2-gagmod.BW965
; US-10-190-435-36

Query Match      99.9%; Score 1477.4; DB 15; Length 4615;
Best Local Similarity 99.9%; Pred. No. 8.7e-308;
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATGGGCGCCGCGCCAGCATCTGCGGCGGCGAGCTGAGCGCTTGGAGCGCATTCGC 60
DB      3137 ATGGGCGCCGCGCCAGCATCTGCGGCGGCGAGCTGAGCGCTTGGAGCGCATTCGC 3196
QY      61 CTGCGCCCGCGCGCAAGAGTCTATCATATGAAGCACTTGTGTGGGCGACCCGCGAG 120
DB      3197 CTGCGCGCCGCGCGCAAGAGTCTATCATATGAAGCACTTGTGTGGGCGACCCGCGAG 3256
QY      121 CTGAGAGAGTTGCGCCCTGAAACCCCGGCTGTGAGAGCAAGGAGGCTGCAAGAGATC 180
DB      3257 CTGAGAGAGTTGCGCCCTGAAACCCCGGCTGTGAGAGCAAGGAGGCTGCAAGAGATC 3316
QY      181 ATCCGCAAGCTGACCCCGGCTGACAGCCGCGAGCGAGAGCTGAAGAGCTTTCAAC 240
DB      3317 ATCCGCAAGCTGACCCCGGCTGACAGCCGCGAGCGAGAGCTGAAGAGCTTTCAAC 3376
QY      241 ACCGTGGCCCACTCTGTAATCTGCTGTCAGAGAGAGATGAGTCCGCGCAACCAAGAGGCC 300
DB      3377 ACCGTGGCCCACTCTGTAATCTGCTGTCAGAGAGAGATGAGTCCGCGCAACCAAGAGGCC 3436
QY      301 CTGGAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB      3437 CTGGAACAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3496
QY      361 GCCGACAAGGCGAAGGTGAGCCAGAACTACCCCATCTGTGCAAGAACTTGCAGAGGCGCAGATG 420
DB      3497 GCCGACAAGGCGAAGGTGAGCCAGAACTACCCCATCTGTGCAAGAACTTGCAGAGGCGCAGATG 3556
QY      421 GTGCAACCAAGGCTACAGCCCGCGCACTTGAACGCTGTGGTGAAGGTGATGAGAGAGAG 480
DB      3557 GTGCAACCAAGGCTACAGCCCGCGCACTTGAACGCTGTGGTGAAGGTGATGAGAGAGAG 3616
QY      481 GCGTTAGCGCCGAGGATGATCCCATGTTTCAACGCGCTGAGCGAGGCGCGCACCCCGAG 540
DB      3617 GCGTTAGCGCCGAGGATGATCCCATGTTTCAACGCGCTGAGCGAGGCGCGCACCCCGAG 3676
QY      541 GACCTGAACAGATGTTGAACACCTGTGGGCGGCGCACAGGCGCGCATGAGATGCTGAAG 600
DB      3677 GACCTGAACAGATGTTGAACACCTGTGGGCGGCGCACAGGCGCGCATGAGATGCTGAAG 3736
QY      601 GACACATCAACAGAGAGGCGCGCGAGTGGAGACCGGCTGCAACCCCGTGCACGCGCGGCC 660
DB      3737 GACACATCAACAGAGAGGCGCGCGAGTGGAGACCGGCTGCAACCCCGTGCACGCGCGGCC 3796
QY      661 ATGCGCCCGCGCGAGATGCGCGAGGCGCGCGGCGAGAGAGATGCGCGCGCACCAAGAGACC 720
DB      3797 ATGCGCCCGCGCGAGATGCGCGAGGCGCGCGGCGAGAGAGATGCGCGCGCACCAAGAGACC 3856
QY      721 CTGCAAGAGAGAGATGCGCTGTGATGACCAAGCAACCCCGCATCCCGTGGGCGCATCTAC 780
DB      3857 CTGCAAGAGAGAGATGCGCTGTGATGACCAAGCAACCCCGCATCCCGTGGGCGCATCTAC 3916
```

```

QY      781 AAGCGGTGATCATCTCTGGGCGCTGAACAAGATGTGCGGATGTATACAGCCCGTGAGCATC 840
DB      3917 AAGCGGTGATCATCTCTGGGCGCTGAACAAGATGTGCGGATGTATACAGCCCGTGAGCATC 3976
QY      841 CTGGAATCAAGAGAGGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB      3977 CTGGAATCAAGAGAGGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4036
QY      901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB      4037 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4096
QY      961 GTGCAAGAGCGCAACCCCGAGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB      4097 GTGCAAGAGCGCAACCCCGAGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4156
QY      1021 CTGAGAGAGATGATGACCGCTGTCAGAGGCGGTGGGCGGCGCCCAAGCAAGAGCGCGCTG 1080
DB      4157 CTGAGAGAGATGATGACCGCTGTCAGAGGCGGTGGGCGGCGCCCAAGCAAGAGCGCGCTG 4216
QY      1081 CTGCGCGAGCGAGATGAGCGCAAGCGCAACCAAGCTGATGATGCAAGAGCAACTTCAAG 1140
DB      4217 CTGCGCGAGCGAGATGAGCGCAAGCGCAACCAAGCTGATGATGCAAGAGCAACTTCAAG 4276
QY      1141 GCGCCCGCGCGCATCTGTCAGAGTGTCTCAACTGTGCGCAAGAGAGGCGCAATCGCCGCAAC 1200
DB      4277 GCGCCCGCGCGCATCTGTCAGAGTGTCTCAACTGTGCGCAAGAGAGGCGCAATCGCCGCAAC 4336
QY      1201 TGCCGCGCGCCCGCAAGAGAGGCTGCTGAAGTGTGCGCAAGAGAGGCGCAAGATGAAG 1260
DB      4337 TGCCGCGCGCCCGCAAGAGAGGCTGCTGAAGTGTGCGCAAGAGAGGCGCAAGATGAAG 4396
QY      1261 GACTGCAACGAGGCGCAAGGCGCAACTTCTGTGGGAGAGATCTGGCCCAAGAGAGGCGCG 1320
DB      4397 GACTGCAACGAGGCGCAAGGCGCAACTTCTGTGGGAGAGATCTGGCCCAAGAGAGGCGCG 4456
QY      1321 CCGGCAACTTCTGTGAGAGCGCGCGCGAGCCGCGCGCGCGCGAGAGCTTCGCG 1380
DB      4457 CCGGCAACTTCTGTGAGAGCGCGCGCGAGCCGCGCGCGCGCGAGAGCTTCGCG 4516
QY      1381 TTGAGAGAGACACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB      4517 TTGAGAGAGACACCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4576
QY      1441 CTGAAGAGCTGTTGCGCAACGACCCCTGAGCCACTAA 1479
DB      4577 CTGAAGAGCTGTTGCGCAACGACCCCTGAGCCACTAA 4615

RESULT 10
US-10-190-435-38
; Sequence 38, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Batrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 4702
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: sp16mod.TV1-gagmod.BW965
```

US-10-190-435-38

Query Match 99.9%; Score 1477.4; DB 15; Length 4702;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGGGCGCCGCGCGCATCTGCGCGCGCGCATGAGCGCCCTGGAGCGCATCCGC 60  
Db ATGGGCGCCGCGCGCGCATCTGCGCGCGCGCATGAGCGCCCTGGAGCGCATCCGC 3283  
61 CTGGCGCCGCGCGCGCAAGAGTCTACATGATGAGCACTGTGTGTGGCCAGCCGCGAG 120  
Db CTGGCGCCGCGCGCGCAAGAGTCTACATGATGAGCACTGTGTGTGGCCAGCCGCGAG 3343  
121 CTGGAGAAATTTGGCCCTTGAACCCCGGCGCTGTGGAGAACAGAGGGCTTGCAGCATG 180  
Db CTGGAGAAATTTGGCCCTTGAACCCCGGCGCTGTGGAGAACAGAGGGCTTGCAGCATG 3403  
181 ATCCGCGACGTGCACCCCGCGCTGCAGACCGCGCGAGGAGGCTGAAAGAGCTTTCAAC 240  
Db ATCCGCGACGTGCACCCCGCGCGCTGCAGACCGCGCGAGGAGGCTGAAAGAGCTTTCAAC 3463  
241 ACCGTGGCCACCTCTGTACTGCGTGCAGAGAAAGTGAAGTCCGCGACACCAAGAGGCC 300  
Db ACCGTGGCCACCTCTGTACTGCGTGCAGAGAAAGTGAAGTCCGCGACACCAAGAGGCC 3523  
301 CTGGACAAGATGAGAGAGAGAGAGCAAGTGCAGAGCAAGATTCAGAGGCCAGAGGCC 360  
Db CTGGACAAGATGAGAGAGAGAGAGCAAGTGCAGAGCAAGATTCAGAGGCCAGAGGCC 3583  
3524 CTGGACAAGATGAGAGAGAGAGAGCAAGTGCAGAGCAAGATTCAGAGGCCAGAGGCC 3583  
361 GCCGACAAAGGAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Db GCCGACAAAGGAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3643  
3584 GCCGACAAAGGAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3643  
421 GTGGACACAGGCGCATCAGCCCGCGCATCTGAAAGCGCTGGGTGAAGTGAAGAGAGAG 480  
Db GTGGACACAGGCGCATCAGCCCGCGCATCTGAAAGCGCTGGGTGAAGTGAAGAGAGAG 3703  
3644 GTGGACACAGGCGCATCAGCCCGCGCATCTGAAAGCGCTGGGTGAAGTGAAGAGAGAG 3703  
481 GCGTTACGCGCGAGAGTGAATCCCATGTTCAACCGCGCTGAAGAGAGAGAGAGAGAGAG 540  
Db GCGTTACGCGCGAGAGTGAATCCCATGTTCAACCGCGCTGAAGAGAGAGAGAGAGAGAG 3763  
3704 GCGTTACGCGCGAGAGTGAATCCCATGTTCAACCGCGCTGAAGAGAGAGAGAGAGAGAG 3763  
541 GACCTGAACAGATGTTGAACAACCTGTGGCGCGCACAGGCGCGCATGAGATGCTGAAG 600  
Db GACCTGAACAGATGTTGAACAACCTGTGGCGCGCACAGGCGCGCATGAGATGCTGAAG 3823  
3764 GACCTGAACAGATGTTGAACAACCTGTGGCGCGCACAGGCGCGCATGAGATGCTGAAG 3823  
601 GACACCATCAACGAGAGAGCGCGCGAGTGGAGACCGCGTGCACCCGCGCGAGCCGCC 660  
Db GACACCATCAACGAGAGAGCGCGCGAGTGGAGACCGCGTGCACCCGCGCGAGCCGCC 3883  
3824 GACACCATCAACGAGAGAGCGCGCGAGTGGAGACCGCGTGCACCCGCGCGAGCCGCC 3883  
661 ATCGCGCGCGCGAGATGCGCGAGCGCGCGCGAGAGAGATGCGCGCGCGCACACAGCACCC 720  
Db ATCGCGCGCGCGAGATGCGCGAGCGCGCGCGAGAGAGATGCGCGCGCGCACACAGCACCC 3943  
3884 ATCGCGCGCGCGAGATGCGCGAGCGCGCGCGAGAGAGATGCGCGCGCGCACACAGCACCC 3943  
721 CTGGAGAGAGAGATGCGCTGTGATGACAGCAACCCCGCATCCCTGTGGCGCATCTTAC 780  
Db CTGGAGAGAGAGATGCGCTGTGATGACAGCAACCCCGCATCCCTGTGGCGCATCTTAC 4003  
3944 CTGGAGAGAGAGATGCGCTGTGATGACAGCAACCCCGCATCCCTGTGGCGCATCTTAC 4003  
781 AAGCGGTGATCATCTGTGGCGCTGAAACAAGATGTCGGATGTACAGCCCGCTGAGCATC 840  
Db AAGCGGTGATCATCTGTGGCGCTGAAACAAGATGTCGGATGTACAGCCCGCTGAGCATC 4063  
4004 AAGCGGTGATCATCTGTGGCGCTGAAACAAGATGTCGGATGTACAGCCCGCTGAGCATC 4063  
841 CTGGACATCAAGAGAGAGCGCGCAAGAGCGCTTCCGCGCATAGTGGAGCGCTTTCTTCAAG 900  
Db CTGGACATCAAGAGAGAGCGCGCAAGAGCGCTTCCGCGCATAGTGGAGCGCTTTCTTCAAG 4123  
4064 CTGGACATCAAGAGAGAGCGCGCAAGAGCGCTTCCGCGCATAGTGGAGCGCTTTCTTCAAG 4123  
901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4183  
4124 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4183  
961 GTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db GTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4243  
4184 GTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4243

1021 CTGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
Db CTGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4303  
4244 CTGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4303  
1081 CTGGCGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
Db CTGGCGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4363  
4304 CTGGCGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4363  
1141 GGGCGCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200  
Db GGGCGCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4423  
4364 GGGCGCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4423  
1201 TGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
Db TGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4483  
4424 TGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4483  
1261 GACTGACCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db GACTGACCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4543  
4484 GACTGACCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4543  
1321 CCGGCGAACTTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
Db CCGGCGAACTTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4603  
4544 CCGGCGAACTTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4603  
1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Db TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4663  
4604 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4663  
1441 CTGAAGAGCGCTGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
Db CTGAAGAGCGCTGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4702  
4664 CTGAAGAGCGCTGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4702

RESULT 11  
US-10-190-435-17  
; Sequence 17, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Esterlita J.  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 4716  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; ;  
; OTHER INFORMATION: GagProteinMutatRevief\_C  
US-10-190-435-17

Query Match 99.9%; Score 1477.4; DB 15; Length 4716;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGGGCGCCGCGCGCATCTGCGCGCGCGCAAGCTGAGACGCTGGAGCGCATCCGC 60  
Db ATGGGCGCCGCGCGCGCATCTGCGCGCGCGCAAGCTGAGACGCTGGAGCGCATCCGC 66  
7 ATGGGCGCCGCGCGCGCATCTGCGCGCGCGCAAGCTGAGACGCTGGAGCGCATCCGC 66  
Db ATGGGCGCCGCGCGCGCATCTGCGCGCGCGCAAGCTGAGACGCTGGAGCGCATCCGC 66  
61 CTGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
Db CTGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126  
67 CTGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126  
Db CTGGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
121 CTGGAGAAATTTGCGCCTGAAACCCCGCGCTGTGAGAGACCAAGAGAGGCTGCAAGAGATC 180

Db 127 CTGGAGAAAGTTGGCCCTGGAACCCCGGCTGTCTGGAGACGACGAGGGCTTCAAGCAGATC 186  
Qy 181 ATCCGCAAGTGTGACCCCGGCTGTGACAGCCGCAAGGAGAGCTTGAAGACCTTGTTCAC 240  
Db 187 ATCCGCAAGTGTGACCCCGGCTGTGACAGCCGCAAGGAGAGCTTGAAGACCTTGTTCAC 246  
Qy 241 ACCGTGGCAACCTGTACTGCGTGCAGAGAAATGAGAGTCCGCGACCAAGAGAGGC 300  
Db 247 ACCGTGGCAACCTGTACTGCGTGCAGAGAAATGAGAGTCCGCGACCAAGAGAGGC 306  
Qy 301 CTGGACAAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 307 CTGGACAAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
Qy 361 GCCGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Db 367 GCCGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426  
Qy 421 GTGGACCAAGGCGCATCAGCCCGGCGACCTGGAAGCGCTGGGTGAAGGTGATCGAGAGAG 480  
Db 427 GTGGACCAAGGCGCATCAGCCCGGCGACCTGGAAGCGCTGGGTGAAGGTGATCGAGAGAG 486  
Qy 481 GCGTTCAAGCCCGAGAGGTGATCCCGCATGTTCAACCGCCCTGAGCGAGAGAGAGAGAGAG 540  
Db 487 GCGTTCAAGCCCGAGAGGTGATCCCGCATGTTCAACCGCCCTGAGCGAGAGAGAGAGAGAG 546  
Qy 541 GACCTGAACACGATGTTGAACACCGTGGGCGGCGACCAAGGCGCGCATGCAATGCTGAAG 600  
Db 547 GACCTGAACACGATGTTGAACACCGTGGGCGGCGACCAAGGCGCGCATGCAATGCTGAAG 606  
Qy 601 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Db 607 GACACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 666  
Qy 661 ATGCGCCCGGCGAGATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
Db 667 ATGCGCCCGGCGAGATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726  
Qy 721 CTGGACAGAGACGATGCGCTGTGATGACGACGACGACGACGACGACGACGACGACGACG 780  
Db 727 CTGGACAGAGACGATGCGCTGTGATGACGACGACGACGACGACGACGACGACGACGACG 786  
Qy 781 AAGCGGTGATCATCTGTGGGCGTGAACAGATCGTTCGATGAGAGAGAGAGAGAGAGAGAG 840  
Db 787 AAGCGGTGATCATCTGTGGGCGTGAACAGATCGTTCGATGAGAGAGAGAGAGAGAGAGAG 846  
Qy 841 CTGGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 847 CTGGACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906  
Qy 901 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 907 ACCCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966  
Qy 961 GTGGAGAAAGGCAACCCCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 967 GTGGAGAAAGGCAACCCCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026  
Qy 1021 CTGGAGAGAGATGAGACCGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
Db 1027 CTGGAGAGAGATGAGACCGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086  
Qy 1081 CTGGCGCAAGGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
Db 1087 CTGGCGCAAGGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146  
Qy 1141 GGGCGCGCGCGAGATGCTCAAGTGTCTCAAGTGTGCGGAAGAGAGAGAGAGAGAGAGAGAG 1200  
Db 1147 GGGCGCGCGCGAGATGCTCAAGTGTCTCAAGTGTGCGGAAGAGAGAGAGAGAGAGAGAGAG 1206  
Qy 1201 TGCCTGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

Db 1207 TGCCTGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266  
Qy 1261 GACTGACCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db 1267 GACTGACCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326  
Qy 1321 CCCGCACTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
Db 1327 CCCGCACTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1386  
Qy 1381 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Db 1387 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
Qy 1441 CTGAAGAGCTGTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479  
Db 1447 CTGAAGAGCTGTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1485

RESULT 12  
US-10-190-305A-13  
; Sequence 13, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MESEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190,305A  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 4716  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
US-10-190-305A-13

Query Match 99.9%; Score 1477.4; DB 16; Length 4716;  
Best Local Similarity 99.9%; Pred. No. 8.7e-308;  
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGGCGCGCGCGAGAGATCTGTGCGGCGGCAAGCTGAGCGCTGGAGGCGATCCGC 60  
Db 7 ATGGGCGCGCGCGAGAGATCTGTGCGGCGGCAAGCTGAGCGCTGGAGGCGATCCGC 66  
Qy 61 CTGGCGCGCGCGGCAAGAGTGTCAATGATGAAGACCTGTGTGGCGCAGCGCGAG 120  
Db 67 CTGGCGCGCGCGGCAAGAGTGTCAATGATGAAGACCTGTGTGGCGCAGCGCGAG 126  
Qy 121 CTGGAGAAAGTGGCTTGAACCCCGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
Db 127 CTGGAGAAAGTGGCTTGAACCCCGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186  
Qy 181 ATCCGCAAGTGTGACCCCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 187 ATCCGCAAGTGTGACCCCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
Qy 241 ACCGTGGCAACCTGTACTGCGTGCAGAGAAATGAGAGTCCGCGACCAAGAGAGGC 300  
Db 247 ACCGTGGCAACCTGTACTGCGTGCAGAGAAATGAGAGTCCGCGACCAAGAGAGGC 306  
Qy 301 CTGGACAAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 307 CTGGACAAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
Qy 361 GCCGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

```
Db 367 GCCGACAAAGGAGGATGAGCCAGAACCTACCTGTCGAGAACTGTGAGGCGCAGATG 426
Qy 421 GTGCAACAGGCGCATGAGCCCGGCACTTGAAAGCTGTGGTGAAGTGTGTCAGAGAAAG 480
Db 427 GTGCAACAGGCGCATGAGCCCGGCACTTGAAAGCTGTGGTGAAGTGTGTCAGAGAAAG 486
Qy 481 GGCCTCAGGCGCGAGGATGATCCCGATGTTCAACCGGCTGAGAGAGGAGGCGCCACCCCGG 540
Db 487 GCTTCAGGCGCGAGGATGATCCCGATGTTCAACCGGCTGAGAGAGGAGGCGCCACCCCGG 546
Qy 541 GACCTGAACAGATGTTGAACAACCGTGGGCGGCAACAGAGCGGCACTGAGATGCTGAAG 600
Db 547 GACCTGAACAGATGTTGAACAACCGTGGGCGGCAACAGAGCGGCACTGAGATGCTGAAG 606
Qy 601 GACACCATCAAGAGAGGCGCGGAGTGGAGCCGCTGACCCCTGTGACCGCCGCGCC 660
Db 607 GACACCATCAAGAGAGGCGCGGAGTGGAGCCGCTGACCCCTGTGACCGCCGCGCC 666
Qy 661 ATCGCCCGGCGAGATGCGCGAGCGCCGCGGAGGAGATGCGCGGCAACCAAGCACC 720
Db 667 ATCGCCCGGCGAGATGCGCGAGCGCCGCGGAGGAGATGCGCGGCAACCAAGCACC 726
Qy 721 CTGCAAGAGCAGATGCGCTGATGACAGCAACCCCGCATCCCTGTGGCGACATCTAC 780
Db 727 CTGCAAGAGCAGATGCGCTGATGACAGCAACCCCGCATCCCGTGGCGACATCTAC 786
Qy 781 AAGCGGTGATCATCTGGGCTTGAAACAAGATGTCGAGATGTAACGCCCTGTAGCATC 840
Db 787 AAGCGGTGATCATCTGGGCTTGAAACAAGATGTCGAGATGTAACGCCCTGTAGCATC 846
Qy 841 CTGGAATCAAGCAGGAGCGCCCAAGAGCGCTTCCGCGCACTAGTGGACCGCTTCTTCAAG 900
Db 847 CTGGAATCAAGCAGGAGCGCCCAAGAGCGCTTCCGCGCACTAGTGGACCGCTTCTTCAAG 906
Qy 901 ACCCTGCGCGCGAGAGAGCAACCAAGAGTGAAGAACTGATGACGACACCTTCTG 960
Db 907 ACCCTGCGCGCGAGAGAGCAACCAAGAGTGAAGAACTGATGACGACACCTTCTG 966
Qy 961 GTGCAAGAACGCAACCCCGACTGCAAGAACATCTGCGGCGCTCTGGGCGCGCGCGCAGC 1020
Db 967 GTGCAAGAACGCAACCCCGACTGCAAGAACATCTGCGGCGCTCTGGGCGCGCGCGCAGC 1026
Qy 1021 CTGGAAGAGATGATGACGCTGCAAGAGGCGCTGAGGCGCGCCCAAGCAAGGCGCGCTG 1080
Db 1027 CTGGAAGAGATGATGACGCTGCAAGAGGCGCTGAGGCGCGCCCAAGCAAGGCGCGCTG 1086
Qy 1081 CTGGAAGAGGCGATGAGCCAGGCGCAACCAAGCGTGTATGATGCAAGAGCAACTTCAAG 1140
Db 1087 CTGGAAGAGGCGATGAGCCAGGCGCAACCAAGCGTGTATGATGCAAGAGCAACTTCAAG 1146
Qy 1141 GGGCGCGCGCGATGCTCAAGTCTTCAAGCTGCGGCAAGAGGCGCAATGCGCGCGCAAC 1200
Db 1147 GGGCGCGCGCGATGCTCAAGTCTTCAAGCTGCGGCAAGAGGCGCAATGCGCGCGCAAC 1206
Qy 1201 TGCCGCGCGCGCGCAAGAGGCGCTGCTGAAGTGGCGCAAGAGGCGCGCAAGATGAAG 1260
Db 1207 TGCCGCGCGCGCGCAAGAGGCGCTGCTGAAGTGGCGCAAGAGGCGCGCAAGATGAAG 1266
Qy 1261 GACTGCAACGAGGCGCGCAACTTCTGTGGCAAGATGTGGCGCAAGGCGCGC 1320
Db 1267 GACTGCAACGAGGCGCGCAACTTCTGTGGCAAGATGTGGCGCAAGGCGCGC 1326
Qy 1321 CCGGCGCAACTTCTGTGAAGCGCGCGCAAGGCGCAACCGCGCGCGCGAGGCTTCCG 1380
Db 1327 CCGGCGCAACTTCTGTGAAGCGCGCGCAAGGCGCAACCGCGCGCGCGAGGCTTCCG 1386
Qy 1381 TTGCAAGAGCAACCCCGCGCAAGAGCAGAGAGCAAGAGCAAGGCGCGCAAGC 1440
Db 1387 TTGCAAGAGCAACCCCGCGCAAGAGCAGAGAGCAAGAGCAAGGCGCGCAAGC 1446
Qy 1441 CTGAAGAGCTTGTGGCAAGCAAGCCCTGTAGCCAGTAA 1479
Db 1447 CTGAAGAGCTTGTGGCAAGCAAGCCCTGTAGCCAGTAA 1485
```

```
RESULT 13
US-10-190-435-20
; Sequence 20, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 20
; LENGTH: 2742
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagTatRevNef_C
US-10-190-435-20

Query Match          99.8%; Score 1475.8; DB 15; Length 2742;
Best Local Similarity 99.9%; Pred. No. 2e-307;
Matches 1477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGGCGCGCGCGGCGAGATCTGGGCGGCGGCAAGCTGAGCGCTGTGGAGCGCATCCG 60
Db 7 ATGGGCGCGCGCGGCGAGATCTGGGCGGCGGCAAGCTGAGCGCTGTGGAGCGCATCCG 66
Qy 61 CTGCGCGCGCGCGGCGAGAGATGCTACATGATGAAGCACTGTGTGGCGCAGCGCGAG 120
Db 67 CTGCGCGCGCGCGGCGAGAGATGCTACATGATGAAGCACTGTGTGGCGCAGCGCGAG 126
Qy 121 CTGGAAGATGTCGCTTGAACCCCGGCTGCTGAGAGCAAGAGGCTGCAAGCATC 180
Db 127 CTGGAAGATGTCGCTTGAACCCCGGCTGCTGAGAGCAAGAGGCTGCAAGCATC 186
Qy 181 ATCCGCGAGCTGACCCCGCGCTGCAAGCGGCGAGCGAGAGTGAAGAGCTGTTCAC 240
Db 187 ATCCGCGAGCTGACCCCGCGCTGCAAGCGGCGAGCGAGAGTGAAGAGCTGTTCAC 246
Qy 241 ACCGTGGCACCTGTACTGTGCTGCAAGAGATGAGATCCGCGCAACCAAGAGGCG 300
Db 247 ACCGTGGCACCTGTACTGTGCTGCAAGAGATGAGATCCGCGCAACCAAGAGGCG 306
Qy 301 CTGGAACAAGATGAGAGAGAGCAACAAGTCCAGAGCAAGATTCACAGGCGCGAGGCC 360
Db 307 CTGGAACAAGATGAGAGAGAGCAACAAGTCCAGAGCAAGATTCACAGGCGCGAGGCC 366
Qy 361 GCCGCAAGGCGAGAGTGAAGCAAGATCAACCATCTGTGAAGACTGTGAGGAGAG 420
Db 367 GCCGCAAGGCGAGAGTGAAGCAAGATCAACCATCTGTGAAGACTGTGAGGAGAG 426
Qy 421 GTGCAACAGGCGATGAGCGCGCAACCTTGAAGCGCTGTGAAGAGTGAAGAGAG 480
Db 427 GTGCAACAGGCGATGAGCGCGCAACCTTGAAGCGCTGTGAAGAGTGAAGAGAG 486
Qy 481 GCTTCAGCGCGAGAGTATCCCATGTTCAACCGGCTGAGAGAGGCGCGCAACCCCGG 540
Db 487 GCTTCAGCGCGAGAGTATCCCATGTTCAACCGGCTGAGAGAGGCGCGCAACCCCGG 546
Qy 541 GACCTGAACAGATGTTGAACAACCGTGGGCGGCAACAGGCGCGCATGATGCTGAAG 600
Db 547 GACCTGAACAGATGTTGAACAACCGTGGGCGGCAACAGGCGCGCATGATGCTGAAG 606
Qy 601 GACACCATCAAGAGAGGCGCGAGTGGAGCCGCTGACCCCTGTGACCGCGCGCC 660
```

Db 607 GACACCATCAAGAGAGGCGCGGAGTGAGGACCGGCGTGCACCCGCGGACCGCGGCCCC 666  
Qy 661 ATGCCCCCGGCGCGCATGTCGCGGAGCCCCCGGCGGAGGACATCCCGGCGACACACAGACCC 720  
Db 667 ATCGCCCCCGGCGCGCATGTCGCGGAGCCCCCGGCGGAGGACATCCCGGCGACACACAGACCC 726  
Qy 721 CTGCGAGAGAGATGCGCTGAGATGACAGCAACCCCGCATCCCGGCGGCGCATCTAC 780  
Db 727 CTGCGAGAGAGATGCGCTGAGATGACAGCAACCCCGCATCCCGGCGGCGCATCTAC 786  
Qy 781 AAGCGGTGATCATCTCGGCGCTGAAACAAGATGTCGATGTAACAGCCCGTGAACATC 840  
Db 787 AAGCGGTGATCATCTCGGCGCTGAAACAAGATGTCGATGTAACAGCCCGTGAACATC 846  
Qy 841 CTGGAATCAAGAGAGGCGCGCGGAGGAGCCCTTCGCGACTACGTTGAGCCGCTTTCAG 900  
Db 847 CTGGAATCAAGAGAGGCGCGCGGAGGAGCCCTTCGCGACTACGTTGAGCCGCTTTCAG 906  
Qy 901 ACCGTGCGGCGGAGAGAGCAACCGAGAGGTGAAGACTGATGACCGACACCTGCTG 960  
Db 907 ACCGTGCGGCGGAGAGAGCAACCGAGAGGTGAAGACTGATGACCGACACCTGCTG 966  
Qy 961 GTGCAAGAGCGCAACCCCGACTGCAAGACCATCTCGCGCTCTCGGCGCGGCGCGCAGC 1020  
Db 967 GTGCAAGAGCGCAACCCCGACTGCAAGACCATCTCGCGCTCTCGGCGCGGCGCGCAGC 1026  
Qy 1021 CTGAGAGAGATGATGACCGCTGCGAGGCGTGGCGCGCGCGCGCGCGCGCGCGCG 1080  
Db 1027 CTGAGAGAGATGATGACCGCTGCGAGGCGTGGCGCGCGCGCGCGCGCGCGCGCG 1086  
Qy 1081 CTGGCGAGAGGAGATGAGCGAGCGCAACACAGGTATGATGAGAGAGAGCACTTCAAG 1140  
Db 1087 CTGGCGAGAGGAGATGAGCGAGCGCAACACAGGTATGATGAGAGAGAGCACTTCAAG 1146  
Qy 1141 GCGCGCGCGCGCATCTGCAAGTCTTCAACTGCGGAGAGAGGCGCACTCGCGCGCAAC 1200  
Db 1147 GCGCGCGCGCGCATCTGCAAGTCTTCAACTGCGGAGAGAGGCGCACTCGCGCGCAAC 1206  
Qy 1201 TGCCTGCGCGCGCGCGCAAGAGGCTGCTGAGATGCGCGCAAGAGGCGCACTGAGAG 1260  
Db 1207 TGCCTGCGCGCGCGCGCAAGAGGCTGCTGAGATGCGCGCAAGAGGCGCACTGAGAG 1266  
Qy 1261 GACTGACAGAGGCGCGAGCGCACTCTGCGGCAATGCTGCGCGCAAGAGGCGCGCG 1320  
Db 1267 GACTGACAGAGGCGCGAGCGCACTCTGCGGCAATGCTGCGCGCAAGAGGCGCGCG 1326  
Qy 1321 CCGCGCACTTCTCTGAGAGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380  
Db 1327 CCGCGCACTTCTCTGAGAGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386  
Qy 1381 TTGAGAGAGAGCAACCCCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Db 1387 TTGAGAGAGAGCAACCCCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446  
Qy 1441 CTGGAAGAGCTGTTGCGGCAAGCAACCCCGTGAAGCACTAA 1479  
Db 1447 CTGGAAGAGCTGTTGCGGCAAGCAACCCCGTGAAGCACTAA 1485

RESULT 14  
US-10-190-305A-15  
; Sequence 15, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190,305A  
; CURRENT FILING DATE: 2002-07-05  
; NUMBER OF SEQ ID NOS: 93

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 2742  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagRevNet\_C  
US-10-190-305A-15  
Query Match 99.8%; Score 1475.8; DB 16; Length 2742;  
Best Local Similarity 99.9%; Pred. No. 26-307;  
Matches 1477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ATGGGCGCGCGCGCGCGCATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 7 ATGGGCGCGCGCGCGCGCGCATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66  
Qy 61 CTGCGCGCGCGCGCGCAAGAGTCTACATGATGAAAGCACTGAGTGGCGCGCGCGCGAG 120  
Db 67 CTGCGCGCGCGCGCGCAAGAGTCTACATGATGAAAGCACTGAGTGGCGCGCGCGCGAG 126  
Qy 121 CTGAGAGAGTTCGCTGTAACCCCGCGCTGCTGAGAGCAAGCGAGGCTGCAAGCAGATC 180  
Db 127 CTGAGAGAGTTCGCTGTAACCCCGCGCTGCTGAGAGCAAGCGAGGCTGCAAGCAGATC 186  
Qy 181 ATCCGCGAGCTGCAACCCCGCGCTGCGAGACCGCGAGAGAGCTGAAAGCGCTTTCAG 240  
Db 187 ATCCGCGAGCTGCAACCCCGCGCTGCGAGACCGCGAGAGAGCTGAAAGCGCTTTCAG 246  
Qy 241 ACCGTGCGCGCGCTGATCTGCTGTCGAGAGAGATGAGTCCGCGCGCGCGCGCGCGCG 300  
Db 247 ACCGTGCGCGCGCTGATCTGCTGTCGAGAGAGATGAGTCCGCGCGCGCGCGCGCGCG 306  
Qy 301 CTGGAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 307 CTGGAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366  
Qy 361 GCGGCAAGAGGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Db 367 GCGGCAAGAGGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426  
Qy 421 GTGCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Db 427 GTGCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486  
Qy 481 GCGTTGAGCGCGGAGAGTTCCTCAATGTTCAACCGCGCTGAGAGAGAGAGAGAGAG 540  
Db 487 GCGTTGAGCGCGGAGAGTTCCTCAATGTTCAACCGCGCTGAGAGAGAGAGAGAGAG 546  
Qy 541 GACTGAAACGATGTTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
Db 547 GACTGAAACGATGTTGAACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606  
Qy 601 GACACCATCAAGAGAGAGGCGCGGAGTGGAGCGGCGTGCACCCGTCGACCGCGCGCG 660  
Db 607 GACACCATCAAGAGAGAGGCGCGGAGTGGAGCGGCGTGCACCCGTCGACCGCGCGCG 666  
Qy 661 ATCGCGCGCGCGCGAGATGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
Db 667 ATCGCGCGCGCGCGAGATGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726  
Qy 721 CTGCGAGAGAGATGCGCTGAGATGACAGCAACCCCGCATCCCGTGGCGCGCATCTAC 780  
Db 727 CTGCGAGAGAGATGCGCTGAGATGACAGCAACCCCGCATCCCGTGGCGCGCATCTAC 786  
Qy 781 AAGCGGTGATCATCTCGGCGCTGAAACAAGATGTCGATGTAACAGCCCGTGAACATC 840  
Db 787 AAGCGGTGATCATCTCGGCGCTGAAACAAGATGTCGATGTAACAGCCCGTGAACATC 846  
Qy 841 CTGGAATCAAGAGAGGCGCGCGGAGGAGCCCTTCGCGACTACGTTGAGCCGCTTTCAG 900  
Db 847 CTGGAATCAAGAGAGGCGCGCGGAGGAGCCCTTCGCGACTACGTTGAGCCGCTTTCAG 906



Qy	901	ACCTGCGCGCCGAGCAGACACCCAGAGGTGAAGAACTGATATGCCGACACCTCTGTTG	960
Db	907	ACCTGCGCGCGCGAGCAGACGACCCAGAGAGGTGAAGAACTGATATGCCGACACCTCTGTTG	966
Qy	961	GTGCAGAACGCGCAACCCCGGACTGCAGAGACCAATCTCTGGCGCGCTCTGGCGCCGGCGCCAGC	1020
Db	967	GTGCAGAACGCGCAACCCCGGACTGCAGAGACCAATCTCTGGCGCGCTCTTGAGCCCGGCGCCAGC	1026
Qy	1021	CTGGAGAGAGATGATGACCGCCTTGCCAGGGCGTGGCGCGCCCGAGCCACAGAGCCCGCGTGT	1080
Db	1027	CTGGAGAGAGATGATGACCGCCTTGCCAGGGCGTGGCGCGCCCGAGCCACAGAGCCCGCGTGT	1086
Qy	1081	CTGGCGCGAGGGGATGATGAGCCAGGCGCCACACACGCTGTATGATGACAGAGACCACTTCAAG	1140
Db	1087	CTGGCGCGAGGGGATGATGAGCCAGGCGCCACACACGCTGTATGATGACAGAGACCACTTCAAG	1146
Qy	1141	GGCCCCCGCGCGCATGTCAAGTGTCTTCAACTGCGCGCAGAGAGGGGCACATGCGCCGCGCAC	1200
Db	1147	GGCCCCCGCGCGCATGTCAAGTGTCTTCAACTGCGCGCAGAGAGGGGCACATGCGCCGCGCAC	1206
Qy	1201	TGCGCGCGCCCCCGCGAGAGAGGGCTGTGGAAGTGCGGCAGAGAGGGCCACAGATGAG	1260
Db	1207	TGCGCGCGCCCCCGCGAGAGAGGGCTGTGGAAGTGCGGCAGAGAGGGCCACAGATGAG	1266
Qy	1261	GACTGTCACATCCGAGCGCCAGGCCCACTTCTTGCGGCAGAGATCTTGCGCCAGCCACAAGGGCGCG	1320
Db	1267	GACTGTCACATCCGAGCGCCAGGCCCACTTCTTGCGGCAGAGATCTTGCGCCAGCCACAAGGGCGCG	1326
Qy	1321	CCCGGCAACTCTCTGACAGAGCGCGCCGAGGCCACACCGCGCCCGCCCGCGCGAGAGCTTCGCG	1380
Db	1327	CCCGGCAACTCTCTGACAGAGCGCGCCGAGGCCACACCGCGCCCGCCCGCGCGAGAGCTTCGCG	1386
Qy	1381	TTTCGAGAGACCAACCCCGCGCCAGAGACGAGAGACAGAGACCGCGAGACCTTGATCCAGC	1440
Db	1387	TTTCGAGAGACCAACCCCGCGCCAGAGACGAGAGACAGAGACCGCGAGACCTTGATCCAGC	1446
Qy	1441	CTGAAGAGCTGTTCGGCAACGACCCCTTGAGCCACTTA 1479	
Db	1447	CTGAAGAGCTGTTCGGCAACGACCCCTTGAGCCACTTA 1485	

```

RESULT 15
US-10-190-435-9
? Sequence 9, Application US/10190435
? Publication No. US20030143248A1
? GENERAL INFORMATION:
? APPLICANT: ZUR MEGEDE, Jan
? APPLICANT: BARNETT, Susan W.
? APPLICANT: LIAN Ying
? APPLICANT: ENGELBRECHT, Susan
? APPLICANT: VAN KENSBURG, Estrella J.
? TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
? TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
? FILE REFERENCE: EP1813.003 / 2302-18133
? CURRENT APPLICATION NUMBER: US/10/190,435
? CURRENT FILING DATE: 2002-12-30
? NUMBER OF SEQ ID NOS: 319
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 9
? LENGTH: 3930
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: GagCompPolmut_C
US-10-190-435-9

```

	Query March	99.8%	Score 1475.8;	DB 15;	Length 3930;
	Best Local Similarity	99.9%;	Pred. No. 1.9e-307;		
	Matches 1477;	Conservative	0;	Mismatches 2;	Indels 0; Gaps 0;
QY	1 ATGGAGCGCCCGCGCAGCATCTCTGCGCGCGCGCAAGCTGAGACCTCTGGAGAGCGCATCCGC	60			
db	7 ATGGAGCGCCCGCGCAGCATCTCTGCGCGCGCGCAAGCTGAGACCTCTGGAGAGCGCATCCGC	66			

QY	61	CTGGCCCCGGGGCGAAGAATGGTCACTATGTAAGCAACCTGGGTGGGGCCAGCCGGAG	120
Db	67	CTGGCCCCGGGGCGAAGAATGTCTACTATGTAAAGCACTTGGTGGGGCCAGCCGGAG	126
QY	121	CTGGAGAAGTTTCGGCTTGAAACCCCGGCTGTCTGGAGACCAAGCGGGCTGCAACAGATC	180
Db	127	CTGGAGAAGTTTCGGCTTGAAACCCCGGCTGTCTGGAGACCAAGCGGGCTGCAACAGATC	186
QY	181	ATCCGCCAGCTGCACCCCGCTTGCACACCGGACCGAGAGCTGAAGAGCTTGTCAAC	240
Db	187	ATCCGCCAGCTGCACCCCGCTTGCACACCGGACCGAGAGCTGAAGAGCTTGTCAAC	246
QY	241	ACCGTGGCCACCTCTGTACTGCTGTCAAGAGAATTCAGAGTCCGCGCAACCAAGAGGCC	300
Db	247	ACCGTGGCCACCTCTGTACTGCTGTCAAGAGAATTCAGAGTCCGCGCAACCAAGAGGCC	306
QY	301	CTGGACAAGATCGAGAGAGACAGAACAAATGTCCAGACAGAAAGATCCAGAGGCCGAGGCC	360
Db	307	CTGGACAAGAGATCGAGAGAGACAGAACAAATGTCCAGAGAGATCCAGAGGCCGAGGCC	366
QY	361	GGCGACAAGGGCCAAAGGTGAAGCCAAATCAACCCCATCTGGAGAACTCTGAGGGCCAGATG	420
Db	367	GGCGACAAGGGCCAAAGGTGAAGCCAAATCAACCCCATCTGGAGAACTCTGAGGGCCAGATG	426
QY	421	GTGCAACAGGCGCATCAGCCCCCGACCCCTGACACGCTTGGGTGAAGGTATCGAGAGAG	480
Db	427	GTGCAACAGGCGCATCAGCCCCCGACCCCTGACACGCTTGGGTGAAGGTATCGAGAGAG	486
QY	481	GCCTTCAGCCCCCGAGGTGATCCCATATTTCACCGCCCTTGAGCCAGGGCGCACCCCCAG	540
Db	487	GCCTTCAGCCCCCGAGGTGATCCCATATTTCACCGCCCTTGAGCCAGGGCGCACCCCCAG	546
QY	541	GACCTGAACACGATGTTGAACACCGTGGGGCGGCCACAGGCGCGCATGAGATCTCTAAG	600
Db	547	GACCTGAACACGATGTTGAACACCGTGGGGCGGCCACAGGCGCGCATGAGATCTCTAAG	606
QY	601	GACACCATCAACGAGAGAGGCCGCGGAGTGGACCGCGTGCAACCCCGTGCAAGCCGCGCCCC	660
Db	607	GACACCATCAACGAGAGAGGCCGCGGAGTGGACCGCGTGCAACCCCGTGCAAGCCGCGCCCC	666
QY	661	ATCGCCCCCGGGCCAGATGCGCGAACCCTCGGGGACGACATCGCGCGGACCAACAGACACC	720
Db	667	ATCGCCCCCGGGCCAGATGCGCGAACCCTCGGGGACGACATCGCGCGGACCAACAGACACC	726
QY	721	CTGCAGAGGACGATCGGCTTGATACGACGACACCCGCCCATCCCGCTGGGGGACATCTAC	780
Db	727	CTGCAGAGGACGATCGGCTTGATACGACGACACCCGCCCATCCCGCTGGGGGACATCTAC	786
QY	781	AAGGGGTGTGATCATCTCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCGGTGAGATC	840
Db	787	AAGGGGTGTGATCATCTCTGGGCTTGAACAAGATCGTGGGATGTACAGCCCGGTGAGATC	846
QY	841	CTGCACATCAAGCAGGGGCCCCCAAGAGACCCCTTCGCGACCTACGTGACCGCTTTCAAG	900
Db	847	CTGCACATCAAGCAGGGGCCCCCAAGAGACCCCTTCGCGACCTACGTGACCGCTTTCAAG	906
QY	901	ACCTGTGCGCGCCGAGCAGAGCACCCAGAGAGGTGAAGACTTGATGACCGACCCCTGCTG	960
Db	907	ACCTGTGCGCGCCGAGCAGAGCACCCAGAGAGGTGAAGACTTGATGACCGACCCCTGCTG	966
QY	961	GTGCAAGAACGCAACCCCGACTGGAACAACATCTGGGCGCTCTCGGGCCCCCGGGCCAGC	1020
Db	967	GTGCAAGAACGCAACCCCGACTGGAACAACATCTGGGCGCTCTCGGGCCCCCGGGCCAGC	1026
QY	1021	CTGAGAGAGATGATGACCGGCTTGCACAGGGGTGGGGCGGCCCAAGCGCCCGGCTG	1080
Db	1027	CTGAGAGAGATGATGACCGGCTTGCACAGGGGTGGGGCGGCCCAAGCGCCCGGCTG	1086
QY	1081	CTGGCCGAGGCGATGAGCCAGGCCAACACCAAGCTGTATGATGCAGAGAGCACTTCAAG	1140
Db	1087	CTGGCCGAGGCGATGAGCCAGGCCAACACCAAGCTGTATGATGCAGAGAGCACTTCAAG	1146



Qy	1141	GGCCCCCGCGCATCGTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGCAAC	1200
Db	1147	GGCCCCCGCGCATCGTCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGCAAC	1206
Qy	1201	TGCCCCGCCCCCGCAAGAGGGCTGTGAAAGTGGGCAAGAGGGCCACAGATGAAG	1260
Db	1207	TGCCCCGCCCCCGCAAGAGGGCTGTGAAAGTGGGCAAGAGGGCCACAGATGAAG	1266
Qy	1261	GACTGCAACGAGCGCCCAACTTCTGGGCAAGATCTGGGCCAGCCACAAGGGCCGC	1320
Db	1267	GACTGCAACGAGCGCCCAACTTCTGGGCAAGATCTGGGCCAGCCACAAGGGCCGC	1326
Qy	1321	CCCGGCACTTCTGCAAGCGCCCGAGCCCAACCGCCCGCGAGAGCTTCCGC	1380
Db	1327	CCCGGCACTTCTGCAAGCGCCCGAGCCCAACCGCCCGCGAGAGCTTCCGC	1386
Qy	1381	TTGAGAGAGACCAACCCCGGCCAGAAAGAGAGAGCAAGAGCCGCGAGACCTGACCAAC	1440
Db	1387	TTGAGAGAGACCAACCCCGGCCAGAAAGAGAGAGCAAGAGCCGCGAGACCTGACCAAC	1446
Qy	1441	CTGAAGAGCTGTGCGCAACGACCCCTGAGCCAGTAA	1479
Db	1447	CTGAAGAGCTGTGCGCAACGACCCCTGAGCCAGTAA	1485

Search completed: March 13, 2005, 04:29:36  
 Job time : 901.894 secs

THIS PAGE BLANK (USPTO)